VISUALISING ARTICULATION: REAL-TIME ULTRASOUND VISUAL BIOFEEDBACK AND VISUAL ARTICULATORY MODELS AND THEIR USE IN TREATING SPEECH SOUND DISORDERS ASSOCIATED WITH SUBMUCOUS CLEFT PALATE

ZOE ROXBURGH

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

QUEEN MARGARET UNIVERSITY
2018

Abstract

Background: Ultrasound Tongue Imaging (UTI) is growing increasingly popular for assessing and treating Speech Sound Disorders (SSDs) and has more recently been used to qualitatively investigate compensatory articulations in speakers with cleft palate (CP). However, its therapeutic application for speakers with CP remains to be tested. A different set of developments, Visual Articulatory Models (VAMs), provide an offline dynamic model with context for lingual patterns. However, unlike UTI, they do not provide real-time biofeedback. Commercially available VAMs, such as Speech Trainer 3D, are available on iDevices, yet their clinical application remains to be tested.

Aims: This thesis aims to test the diagnostic use of ultrasound, and investigate the effectiveness of both UTI and VAMs for the treatment of SSDs associated with submucous cleft palate (SMCP).

Method: Using a single-subject multiple baseline design, two males with repaired SMCP, Andrew (aged 9;2) and Craig (aged 6;2), received six assessment sessions and two blocks of therapy, following a motor-based therapy approach, using VAMs and UTI. Three methods were used to measure therapy outcomes. Firstly, percent target consonant correct scores, derived from phonetic transcriptions provide outcomes comparable to those used in typical practice. Secondly, a multiple-phonetically trained listener perceptual evaluation, using a two-alternative multiple forced choice design, to measure listener agreement provides a more objective measure. Thirdly, articulatory analysis, using qualitative and quantitative measures provides an additional perspective able to reveal covert errors.

Results and Conclusions: There was overall improvement in the speech for both speakers, with a greater rate of change in therapy block one (VAMs) and listener agreement in the perceptual evaluation. Articulatory analysis supplemented phonetic transcriptions and detected covert articulations and covert contrast as well as supporting the improvements in auditory outcome scores. Both VAMs and UTI show promise as a clinical tool for the treatment of SSDs associated with CP.

Key Words: Cleft Palate, Speech, Ultrasound, Visual Biofeedback, Visual Articulatory Models, Phonetic Transcriptions, Perceptual Evaluation

Acknowledgments

First and foremost, I would like to thank my supervisory team – Prof James, M. Scobbie, Dr Joanne Cleland (of University of Strathclyde) and Dr Sara Wood. I could not have asked for a better team. Without you all this thesis would not have been possible. You have all offered invaluable feedback and support throughout my PhD journey and I could not have done it without either one of you. Not only have you offered professional support, but have offered life experiences and friendships I will never forget. You have supported all of my decisions in my career (even offered me a job) and personal life and for that, I can't thank you enough!

Secondly, I would like to thank my child participants and their families for the huge commitment they made to the project. From an academic and clinical perspective, I learned so much from the time spent together and I loved every minute of working with them. I would also like to thank Gillian Cairns and Natalie Lodge, from the specialist Cleft Lip and Palate Speech and Language Therapy Service in NHS Lothian, for recruitment and input throughout the data collection period. Thank you also to all of the participants who took part in the perceptual evaluation and to Prof James, M. Scobbie and Dr Joanne Cleland for completing inter-rater reliability measures.

I would like to especially thank Steve Cowen for his technical support and ever valued friendship throughout my journey. Thank you for encouraging me to apply for the PhD in the first place, listening to my rants along the way, offering advice, building my confidence and always believing in me.

Thank you to Prof Alan Wrench for his speedy technical assistance and for always taking his time to offer advice and show an interest in my research. I would also like to thank Alan for providing the milk for the ever-needed, endless cups of tea!

Thank you to Robert Rush and Felix Schaeffler for statistical advice, and to Eleanor Lawson for her help with the design of the perceptual evaluation.

Thank you to all of the other members of staff in Speech and Hearing Sciences for your kind support, with special thanks to Prof Janet Beck for welcoming me into the division, and to Sonja Schaeffler for being a great friend. I would like to further my thanks to other members of staff at Queen Margaret University, for always taking the time to ask me about my progress. It has been a pleasure being part of the Speech and Hearing Sciences Division and wider university community. Also thanks to Anne Harding-Bell and Sara Howard, along with the rest of the team, at the University of Sheffield for inviting me to teach on your cleft palate course and for offering encouragement and expert advice on cleft palate throughout my PhD.

Thank you to my internal and external PhD examiners Dr Joan Ma and Prof Rachael-Anne Knight for their time, advice and an enjoyable viva!

I would also like to thank NHS Lothian and NHS Grampian Aberdeenshire Speech and Language Therapists for your support of my research role. I am particularly grateful to Emma-Louise MacPherson and Jean Loggie for making allowances in my clinical post in order for me to complete my thesis over the final few months, your support and encouragement has been invaluable.

I would like to say a massive thank you to my fiancé Andrew, who has had to put with me (and my PhD) and has been my rock throughout the whole process. I also owe the hugest of thanks to my family for being my biggest fans, particularly my parents. They have always understood how important my PhD is to me and have

offered more encouragement than I can ever thank them for. I would like to dedicate my PhD thesis to my Aunty Karen, who suddenly passed away in the final months of my journey. She was always a huge supporter of me and my successes.
Finally, I am thanking all of my friends. I would like to especially thank Liberty, Casey, Emma and Suzanne for always being there and picking me up when I felt down, letting me moan and most importantly celebrating my successes and making me laugh! I really would not have succeeded without my amazing fiancé, family, friends and supervisory team, so once again – thank you all!

Table of Contents

List of Figures

List of Tables

1 Theoretical Background

Children with Cleft Palate (CP) are commonly known to present with speech difficulties. Speech difficulties associated with a palatal cleft may become apparent when babies begin to vocalise prior to surgical management. These speech difficulties may persist, despite an adequately established oral-pharyngeal mechanism post-surgery (Peterson-Falzone et al. 2010).

Speech and Language Therapists (SLTs) working with CP usually use intervention approaches such as: traditional articulation therapy (Van Riper 1978), phonological approaches (Harding-Bell and Howard 2011) and psycholinguistic approaches (Stackhouse and Wells 1997), with articulation therapy preferred for speakers with CP (Peterson-Falzone et al. 2006). Similar to an articulation therapy approach is a motor-based therapy approach (e.g. Preston et al. 2014). When using these approaches, it can be difficult for SLTs to describe to a client how they are moving their inner articulators, such as their tongue, as they are not visible during speech. However, technological advances have made new tools available, which allow therapists to diagnose and describe to clients more easily their articulation problems, and which provide visual biofeedback in therapy. In the broad sense, the term biofeedback refers to individuals learning to self-regulate and positively changing an automatic physiological function of which they would not otherwise be aware of, through monitoring (France and DeAngelo 2016). This is not necessarily the case when self-monitoring and changing speech patterns as we, as speakers, have conscious control over our articulators, even though these are largely hidden to the human eye.

One such visual biofeedback method, electropalatography (EPG), is recommended by the Royal College of Speech and Language Therapists (RCSLT) as a treatment option for school-aged children with persistent articulatory disorders associated with CP (RCSLT 2005). EPG indirectly shows the effects of the cleft using a standardised palate for those with typical, or in the case of CP, atypical, palate shapes and sizes. Similarly, ultrasound tongue imaging (UTI) also shows the indirect effect of the cleft on speech. As UTI shows the surface of the tongue with most of the tip to the root being visible, it is particularly useful for looking at backing (Bressmann et al. 2011), which is a common compensatory articulation in the speech of individuals with CP (Harding and Grunwell, 1998; Sell et al. 1999; Peterson-Falzone et al. 2010). Two studies (Gibbon and Wolters 2005; Bressmann et al. 2011) have explored the compensatory articulations in the speech of individuals with CP using UTI and Zharkova (2013) proposes quantitative measures for analysing ultrasound data in speakers with CP; however, its therapeutic applications remain to be tested.

Quite a different set of developments, namely the recent advances in portable multimedia consumer technology, have enabled developers to produce commercial software aimed at enhancing formal and informal therapeutic intervention in speech disorders, for example apps (applications) for the iPad (Apple 2012). One example, Speech Trainer 3D (Smarty Ears 2011) provides an animated Visual Articulatory Model (VAM) which is intended to be used as a tool to explain articulatory features to clients. VAMs provide an off-line articulatory model for demonstrating lingual movements in relation to the passive articulators, providing a context for lingual movements, unlike ultrasound. However, they do not allow the client to view their own tongue at all, particularly in real-time or offer biofeedback as UTI does. Whilst EPG and ultrasound have been tested clinically for either CP or other populations, VAMs remain to be tested. This thesis aims to investigate the use of ultrasound for diagnostic purposes, comparing articulatory analysis to perceptual assessment using phonetic transcription and multi-listener judgements; and to test the therapeutic application of both UTI and VAMs, to determine whether listeners are able to improve on therapy targets by using an off-line VAM or whether they require the real-time biofeedback provided by ultrasound. Three sets of data will be presented: phonetic transcriptions (including percent target consonant correct scores); a multiple phonetically-trained listener perceptual evaluation and an articulatory analysis of ultrasound data.

The remainder of this chapter provides a synthesis of the background literature pertinent to this study. First of all, it gives an overview of cleft lip and palate, followed by the characteristics of speech in individuals with CP and methods for assessing and treating speech sound disorders associated with cleft palate. The final section of this chapter provides a critical analysis of visual feedback and biofeedback techniques and the clinical applications of these tools, with a particular focus on their application for speakers with cleft palate.

1.1 Cleft Lip and Palate Overview

Orofacial clefts are structural disorders which occur early in the embryo and are therefore present at birth (Peterson-Falzone et al. 2010), caused by a failure of fusion during embryology (Watson 2001; Rahimov et al. 2012). Orofacial clefts can be divided into two groups (Mossey et al. 2009; Peterson-Falzone et al. 2010):

1) Cleft Lip (CL) with or without (+/-) Palate (P);
2) Isolated Cleft Palate / Cleft Palate Only (CP/ CPO)

1.1.1 Incidence and Prevalence of Cleft Lip and Palate

Collectively, orofacial clefts are the second most common birth defect (Levi et al. 2011) and are specifically the most common craniofacial condition (*CSAG*; Sandy et al. 1998; Rahimov, et al. 2012). In 2009 the incidence was around 1.7 per 1000 live born babies (Mossey et al. 2009), however more recent studies have shown an average prevalence of approximately 1.2/1000 live births worldwide (Rahimov et al. 2012). Levi et al. (2011) note a range in incidence from 1 in 500 to around 1 in 2,500 births.

Bellis and Wohlgemuth (1999) reported the incidence of infants born with Cleft Lip and Palate (CLP) within the Edinburgh Cleft Units catchment area between January 1971 and December 1990 as 1.4 per 1000 live births (1 in 711). They found that 25% of clefts affected the primary palate, 45% affected the secondary palate and 30% affected both the primary and secondary palates (see Figure 1 below). They found an overall higher percentage of affected males to females (54% male to 42% females), with a higher percentage (56%) of clefts of the secondary palate in females (Bellis and Wolhgemuth 1999). Paterson et al. (2011) investigated the proportion of children with CL and/or P diagnosed prenatally between 1999 and 2008 in those referred for treatment at the Royal Hospital for Sick Children in Glasgow. The percentage of all clefts diagnosed prenatally within their study was 15% which increased to 28% when only CL+/-P was considered (Paterson et al. 2011). They also discovered an increase in prenatal detection between 1999 (11%) and 2008 (50%). Paterson et al. also note variation in the number of cleft cases within the UK.

The National Managed Clinical Network for Cleft Lip and Palate Service in Scotland (Cleft Care Scotland 2016) report the total number of Cleft Births in Scotland between 2000 and 2016 as 1435 births (including 30 deceased patients). Between April 2015 and April 2016, it is reported that there were 91 births in total, with cleft palate (CP) being the cleft type with highest numbers (37 births). There were no children born with bilateral cleft lip (BCL), six born with bilateral cleft lip and palate (BCLP), 23 born with unilateral cleft lip (UCL) and 25 born with unilateral cleft lip and palate (UCLP) (see section 1.1.3 for cleft types).

1.1.2 Embryological Development and Anatomy

Embryological development of the face and primary and secondary palates occurs between weeks four to 12 of gestation (Peterson-Falzone et al. 2010; Rahimov et al. 2012). The primary palate can be described as pre-palatal structures, including the upper lip, alveolar ridge, anterior portion of the maxilla, back to the incisive foramen (Peterson-Falzone et al. 2010; Levi et al. 2011). This forms between weeks five and seven of gestation. The secondary palate extends posteriorly from the incisive foramen, through the hard palate, velum and uvula, and forms between weeks eight and 12 (Levi et al. 2011) (see Figure 1). Fusion of the primary and secondary palates at the incisive foramen, and with the nasal septum, forms the definitive palate (Levi et al. 2011). A palatal cleft can occur anywhere along the Y-shaped lines of fusion, at any stage of development (Atkinson and Howard 2011).

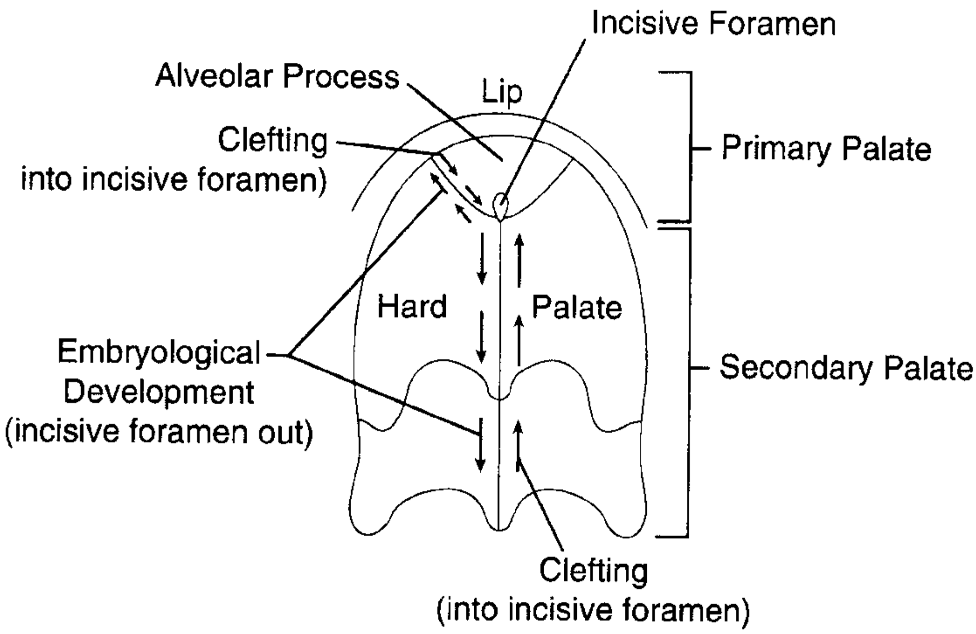


Figure 1 Primary and Secondary Palates (cf. Peterson-Falzone et al. 2010)

1.1.3 Cleft Types and Classifications

Generally, cleft lip (CL) is regarded distinctly different from cleft lip with or without cleft palate (CL+/-P) (Harville et al. 2005). Classification of clefts is of particular importance to provide a basis for research, whether it is epidemiologic, fundamental or clinical (Luijsterburg and Vermeij-Keers 2011).

Within CL+/-P or Cleft Palate Only (CPO), there are various types of cleft with varying levels of severity (see Figure 2). Clefts of the primary palate can be further classified as complete or incomplete, unilateral or bilateral. Unilateral cleft lip is most commonly found on the left side (Kummer 2014). As with clefts of the primary palate, those occurring in the secondary palate (CPO) can also be further classified into complete or incomplete. However, the distinction between unilateral and bilateral clefts of the secondary palate is not always made (Kummer 2001). Clefts of both the primary and secondary palate (CL+/-P) are also common, with a further classification of unilateral, bilateral, complete or incomplete (Kummer 2001; Kummer 2014).

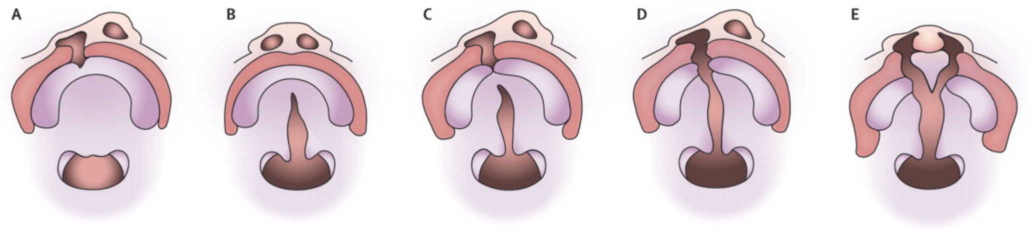


Figure 2 Orofacial Clefts

(A) Cleft lip and alveolus. (B) Cleft palate. (C) Incomplete unilateral cleft lip and palate. (D) Complete unilateral cleft lip and palate. (E) Complete bilateral cleft lip and palate. (Shaw WC. 1993. cf. Mossey, Little, Munger, Dixon and Shaw, 2009)

1.1.3.1 Submucous Cleft Palate

In the absence of an opening into the nasal cavity, there may also be a Submucous Cleft Palate (SMCP) which can be found in the secondary palate (Peterson-Falzone et al. 2010). In the case of SMCP, the oral surface mucosa is intact, however the underlying musculature and structure of the palate is affected (Kummer 2014). The incidence of SMCP ranges between 0.2% and 0.8% of the general population (Stewart et al.1972; Garcia Velasco et al. 1988).

SMCP has three obvious visible stigmata. The first is a bifid uvula, which can be easily viewed during an oral examination. Whilst bifid uvula is commonly associated with SMCP, it is also relatively common to have an isolated bifid uvula with an incidence of around 3% in children within the general population (Shprintzen et al. 1985; Wharton and Mowrer 1992). The second is muscle diastasis or a midline separation of the soft palate musculature which causes the overlying mucosa of the soft palate to look transparent. This is often referred to as a zona pellucida (Kosowski et al. 2012). Finally, the third is a notch into the hard palate (Peterson-Falzone et al. 2010). SMCP may be obvious through an examination of the oral cavity; however, some may only be apparent by carrying out a nasendoscopy to view the nasal side of the velum. They can also range in severity from a bifid uvula, to a cleft under the mucosa lining along the line of fusion to the incisive foramen (Kummer 2014). In around 50% of the cases of a SMCP all three of these classic stigmata will appear together, with around 30-40% of cases where two of these stigmata are present (Miguel et al. 2007). For other patients with SMCP, there are no obvious clinical signs on intraoral examination; however, the muscle diastasis may be identified during imaging or surgery. In these cases, this is known as an occult SMCP (Kaplan 1975). Occult SMCP make up around 10-20% of individuals with SMCP (Miguel et al. 2007).

As the hard and soft palates may appear normal in routine screening, SMCP is often later diagnosed at around ages four or five years (Reiter et al. 2011; Sullivan et al. 2011). Children with SMCPs are often referred to medical services due to speech difficulties associated with Velopharyngeal Dysfunction (VPD), which occur in approximately 5-10% of cases (Gosain et al. 1996; Isotalo et al. 2007). McWilliams (1991) suggests that the primary symptom is hypernasality, secondary to VPD. Section 1.2.2.1.1 below provides a discussion on VPD and speech characteristics associated with VPD are discussed in section 1.2.3. These can include disordered resonance, weak pressure consonants, audible nasal emission and/or nasal turbulence and articulation difficulties. Whilst there are grading systems for rating the severity of a SMPC, it has been recognised that there is no correlation between severity of the anatomical abnormality and the severity of symptoms of VPD (Weatherley-White et al. (1972).

1.1.4 Conditions Associated with Cleft Palate

CLP can also be categorised into syndromic or non-syndromic/isolated clefts. The deciding factor is whether or not there are other physical or developmental disorders present (Meng et al. 2009). The frequency of associated conditions and cleft is reported to be between 20-60% (Stoll et al. 2000), with a higher frequency in clefts of the secondary palate, for example SMCP or occult SCMP (Cohen 1978; Coleman and Sykes 2001).

The majority of CLPs (~70%) occur independently from any other disorders, including other craniofacial abnormalities, and are therefore isolated or non-syndromic (Mossey and Castilla 2003; Levi et al. 2011; Rahimov et al. 2012). For the majority of cases of non-syndromic CLP, it remains difficult to identify specific aetiological influences (Dixon et al. 2011; Levi et al. 2011), however investigations have been carried out into the environmental and genetic risk factors which may be associated with CLP. As there are both genetic and environmental risk factors, isolating a single cause becomes problematic (Levi et al. 2011).

The remaining 30 per cent of clefts are said to be syndromic, with CPO being more commonly associated with congenital malformations (up to 50%) than CL+/-P (around 5-10%) (Mossey and Castilla 2003). Orofacial clefts have been described as part of the disorder in more than 300 syndromes (Meng et al. 2009). More specifically, CL+/-P may be a feature of more than 200 genetic syndromes and CPO is noted as a component of more than 400 genetic disorders (Mossey et al. 2009).

CP can also be part of a sequence or an association if there are multiple anomalies associated with the cleft palate. If it is syndromic, there is a presence of various anomalies due to a single cause such as genetics or teratogens (Dixon et al. 2011). If it is part of a sequence it is likely that multiple anomalies are secondary to a primary anomaly, causing a chain reaction, whereas in an association, the aetiology is unknown (Persson and Sjogreen 2011). Due to the nature of these conditions, speech difficulties may be due to VPD, developmental speech and language disorders, oral motor dysfunction or hearing impairment (Persson and Sjogreen 2011).

Within the hundreds of conditions associated with CP, some of these may also be associated with other structural anomalies. Firstly, unusual shapes or sizes of the pharynx may be associated with conditions in which cleft occurs, for example in 22q11 Deletion Syndrome or Van der Woude Syndrome where the pharynx may be deep or wide; or Treacher Collins syndrome or Apert Syndrome where the pharynx may be narrow. Variations in jaw size may also be associated with conditions linked with cleft, such as in Pierre Robin Sequence. Caouette-Laberge et al. (1994) reported that the prevalence of CP in individuals with Pierre Robin Sequence is around 90%. In the case of SMCP, this can also be associated with a short velum, reduced velar excursion and the failure to close off the oropharynx from the nasopharynx during speech. 22q11 is a common syndrome associated with SMCP, with the prevalence of a SMCP among children with 22q11 deletion syndrome ranging from 21-90% (Golding-Kushner 1985; Shprintzen 2008). Dentition or malocclusion, unusual size of tonsils, hearing impairment and facial asymmetry, for example in Hemi-facial Microsomia (see section 2.2.1), may also have an impact on speech production.

The following section discusses the speech development and characteristics of children born with CP and the active and passive errors associated with VPD, considering the similarities and differences to typically developing children born without CP.

1.2 Speech in Individuals with Cleft Palate

The speech of individuals with CP is impacted by a combination of physical, physiological, cognitive and linguistic factors (Harding and Grunwell 1998). Prior to surgery to close the palatal cleft, there is likely to be hypernasal resonance and difficulties producing consonants requiring intraoral air pressure, due to the coupling between the oral and nasal cavities (Chapman and Willadsen 2011). These errors may persist post-surgery for various reasons. First of all, surgery may require a two-stage procedure, resulting in a hard palate without full closure, or a residual fistula in the hard palate. Secondly, VPD may also occur post-surgery which will impact speech development (Sharp et al. 2003; Chapman and Willadsen 2011). Other factors influencing speech development may include: reduced sensation from scarring, poor dentition and a high incidence of otitis media and conductive hearing loss (Morris and Ozanne 2003). It is reported that otitis media is frequently negatively associated with speech and language development in children with CP (Chapman et al. 2001) and that 90% of children born with CP will already have fluid present in the middle ear (Peterson-Falzone et al. 2010).

Cleft type and severity will also impact speech development in children with CP. It has been reported that children with CL+/-P are more likely to demonstrate more severe articulation difficulties than those with CPO (Peterson-Falzone et al. 2010). This is due to structural conditions such as maxillary collapse, dental malalignment, missing teeth, ectopic eruption of teeth, supernumerary teeth and protrusion of the premaxilla in those with unilateral or bilateral complete clefts (Peterson-Falzone et al. 2010). Riski and DeLong (1984) report that as the severity of the cleft increases, the severity of articulation errors also increase. Table 1 shows a classification of severity in SSDs, based on percent consonant correct (PCC) scores, derived from phonetic transcriptions (Shriberg 1982). This is not, however, specifically for CLP and should therefore be interpreted with caution. Van Lierde et al. (2002) reported no statistical difference in overall speech intelligibility between UCLP and BCLP, however statistical differences were evident between speakers with CP and typically developing speakers.

PCC Scores	Severity Rating
>85%	Mild
65-85%	Mild-Moderate
50-64%	Moderate-Severe
<50%	Severe

Table 1 Shriberg (1982) Severity Ratings for Primary Speech Sound Disorders

Peterson-Falzone et al. (2010) note that individuals with CP are more likely to demonstrate variable and poorer speech skills than children without cleft. Hardin-Jones and Jones (2005) reported that 67% out of 212 preschoolers with CL+/-P were enrolled in or had previously received Speech and Language Therapy (SLT). However, it is expected that the majority of children with repaired CL+/-P will develop acceptable speech skills (Stengelhofen 1989).

Although the speech errors found in children with CP are primarily articulatory in nature, these errors may have a phonological consequence (Harding and Grunwell 1996; Morris and Ozanne 2003). The articulatory and perceptual constraints presented by a cleft can influence a child's phonological development (Harding-Bell and Howard 2011). The following sections describe the development of articulation and phonological processes in children with CP in comparison to typically developing children. Before discussing the cleft type speech characteristics that may occur post-surgery, it is useful to consider early vocalisations in babies.

1.2.1 Pre-Linguistic and Early Speech Development in Children with Cleft Palate

During the first year of life for all children, the biological or physiological constraints of a baby's vocal tract and the language that the baby hears spoken around him/her, along with their linguistic and cognitive skills, are the key factors that influence the production of sounds (Harding and Grunwell 1996; Chapman and Willadson 2011).

As babies' vocal tracts are very small, the larynx sits high, the epiglottis touches the palate and the tongue fills the oral cavity (Harding and Grunwell 1996). Lieberman et al. (1972) describe the vocal tract of a baby as resembling that of a primate. As the tongue is large and the larynx is high and the muscles of the velopharyngeal (VP) mechanism are underdeveloped, a babies breathing is nasal, which in turn leads to early vocalisations being primarily hypernasal and vocalic (Harding and Grunwell 1996; Chapman and Willadson 2011). As the baby reaches around four to six months, oral consonants will become apparent, with velar stops, velar/uvular fricatives and glottal stops being most common (Harding and Grunwell 1996). By six months, oral/nasal and voiced/voiceless contrasts will be evident, glottal productions decrease and supraglottal productions increase (Harding and Grunwell 1996; Chapman and Willadson 2011).

During months five to ten of development canonical babbling emerges (Chapman and Willadsen 2011). During babble, consonants tend to appear from back to front and vowels will appear front to back. Babbling in typically developing children will also contain non-English speech sounds such as bilabial trills, clicks and linguolabials (Harding and Grunwell 1996). By six months, a progression toward anterior placement occurs as babble develops. It is at this stage onward that cleft-type characteristics become evident (Harding and Grunwell 1996).

Although babies with CP will vocalise as much as babies with no cleft, the onset of canonical babbling, acquisition of oral/nasal contrast and the shift from glottal to supraglottal consonants will be disrupted if the cleft is unrepaired or only partially repaired (Chapman and Willadsen, 2011). Lohmander-Agerskov et al. (1994) investigated the babbling patterns of 35 babies with CL+/-P or CPO and two babies with no cleft. Results showed a significantly high frequency (57%) of anterior sounds and low frequency (15%) of posterior sounds in babies with CPO. For children with BCLP or UCLP, there were contrasting findings of a high frequency of posterior sounds, however this was not significant. Children with a cleft in the soft palate only showed complete dominance of anterior placement. The extent of the cleft, cleft width or use of appliance had no significant effect on the manner of articulation. However, plosives were the most commonly used manner of articulation (Lohmander-Agerskov et al. 1994). Chapman et al. (2001) also reported the common use of plosives in their study of 45 babies (30 babies with unrepaired CP and 15 children with no cleft) aged nine months. Other sounds frequently occurring in babbling of babies with cleft were glides, glottals [h] and [ʔ]; and non-English sounds such as [ʁ] and [ɣ] (Chapman et al. 2001). Chapman (1991) calculated the number of vocalisations produced by children with and without CLP and found that those with unrepaired CP overall produced fewer consonants than those children with no cleft. She noted a preference in babies with CP for nasals, glides and glottal fricatives [h]. Hardin-Jones et al. (2003) measured spontaneous speech samples of 53 children (35 with CLP and 18 with CPO). The two groups (CPO and CLP) were compared to detect any differences in canonical babbling, size of consonant inventory, place and manner of articulation and vocalisation frequency. Results showed no significant difference in any of the above; however, babies with CPO produced fewer compensatory stop consonants (i.e. fewer active errors as a result of the cleft, such as retraction, see below for more information) and had a more anterior place of articulation (Hardin-Jones et al. 2003).

Chapman et al. (2003) suggest that frequent use of oral stops in babbling, noted above, will result in better speech and language skills. Lohmander and Persson (2008) suggest that more alveolar productions in babbling will lead to a higher percentage consonant correct (PCC) and that more velar stops in babbling will lead to more retracted productions in later speech. Although the relationship between early speech development and later speech and language in children with CP has been investigated, the evidence is dependent on the outcomes being measured and whether these are measured pre- or post-palatal repair (Chapman and Willadsen 2011).

1.2.2 Speech Outcomes Post-Surgery

Controversy remains with regards to timing and procedures for surgical palatal repair. However, from a speech and language therapist's (SLT's) perspective early, complete closure is preferred (Lohmander 2011) as children begin to show improvement in the development of speech sounds as a result of palatal repair (Chapman et al. 2003; Jones et al. 2003). Chapman et al. (2008) examined the impact of age and lexical status at the time of primary palatal repair in preschool aged children with CP. They suggest that the children who received surgery at a younger age, and were less advanced lexically, achieved better articulation and resonance outcomes at three years of age.

1.2.2.1 Speech Outcomes

Lohmander (2011) carried out a review of the literature based on speech outcomes post-surgery. The articles reviewed were based on assessment from recordings of young children through to young adults, however did not include those articles based on live transcription. Based on her review, Lohmander concluded that the less severe the cleft, the better the outcome. It was also noted that, due to the different surgical techniques and timings employed in the studies, it is difficult to determine what the significant factor for speech outcomes is (Lohmander 2011). Table 2 shows the percentages of acceptable speech post-surgery found in Lohmander's review.

Age Group	Percentage of Acceptable Speech
3 years of age	50-60%
4-5 years of age	60-70%
6-8 years of age	70-80%
10-16 years of age	80%
Young Adults	90-100%

Table 2 Percentages of Acceptable Speech (adapted from Lohmander 2011)

As mentioned above, the majority of children will achieve speech within normal limits after surgery. However, if a two-stage procedure is employed or secondary surgery is required, there may be residual fistulas or VPD which consequently results in cleft-type speech characteristics.

1.2.2.1.1 Velopharyngeal Dysfunction

Velopharyngeal Dysfunction (VPD) is a generic term used to describe a range of disorders resulting from leakage of air into the nasal passage (Woo 2012). Common characteristics of VDP are resonance disorders such as hypernasality and nasal air emission, along with poor intelligibility. Symptoms of VPD may be caused by a number of different factors, for example anatomical abnormalities, such as in the case of a cleft, musculoneuronal causes or mislearned behaviours (Woo 2012). While it may be possible to detect symptoms of VDP associated with CP, such as resonance difficulties, or the cleft-type speech characteristics discussed below, it is not possible to diagnose VPD without instrumental assessment such as nasendoscopy or videofluoroscopy.

As a result of VPD, children may adopt active (compensatory) errors, which can be treated by speech therapy, or passive (obligatory) errors, which will most likely require further surgery. The following section will discuss the cleft-type characteristics associated with CP and VPD, with evidence from perceptual and instrumental assessments.

1.2.3 Cleft-Type Speech Characteristics

Speech disorders associated with CP, and associated VPD, are commonly viewed as articulation disorders (Russell and Grunwell 1993), however Harding and Grunwell (1993; 1996) considered speech characteristics in CP with regards to phonological development. Some children with CP will adopt *passive* or *obligatory* errors and will use only sounds that are readily available due to the constraints of their impaired structure and function of the palate and VP mechanism.

Children using these passive errors tend to have a limited inventory of sonorants [m n ŋ w l j] and the glottal fricative [h] (Harding and Grunwell 1996). Weak articulations, passive nasal fricatives, nasal realisations of plosives and nasal emission accompanying consonants, are also described as passive errors (Harding and Grunwell 1998). As the obligatory errors described are due to an anatomical impairment, they are not easily treated in speech therapy (Golding-Kushner 1995) and, therefore, require further surgery.

In other cases, *active* or *compensatory* strategies are adopted by children with CP in order to facilitate their phonological development and to extend their phonological inventory. These tend to be non-English speech sounds and the use of these strategies is often effective in establishing contrasts between phonemes (Harding and Grunwell, 1996).

Harding and Grunwell (1998) collected and phonetically transcribed speech samples from younger and older speakers with CP. They looked at speech parameters including nasal resonance, turbulence and emission, voice, lip posture, intelligibility and consonant production. Harding and Grunwell (1993; 1998) discuss the active articulations, which are categorised as Cleft-Type Realisations (CTRs). These CTRs include: active nasal fricatives, glottal articulations, pharyngeal fricatives, backing, double articulations, palatal fricatives, lateral fricatives, gliding of fricatives, and imprecise tongue movements, such as tongue tip/blade distortions (Harding and Grunwell 1996; 1998).

Similarly, Sell et al. (1999) propose 10 cleft-type characteristics (CTCs), which they separate into active and passive errors. The first group of errors described by Sell et al. are classified as errors which are actively produced as a substitution for target consonants. These include dentalisation, lateralisation, double articulations, retraction (to velar or uvular position), pharyngeal articulations, glottal articulations (including glottal replacement or glottal reinforcement) and active nasal fricatives. These errors will be discussed in more detail, in line with instrumental analysis, below in sub-section 1.5.1.1. CTCs eight to 10 are passive errors, which occur as a consequence of VPD, or fistulae. These include weak or nasalised consonants, nasal realisation of plosives or nasal realisation of fricatives, which differ from active nasal fricatives in terms of the manner in which they are produced. Active nasal fricatives occur when production of the target /s/ involves active inhibition of oral airflow, whereas nasal realisation of fricatives involves the passive escape of air nasally. This distinction is crucial diagnostically. One method of assessing the distinction is nose-holding. If [s] can be produced with nose holding but is realised nasally without nose-holding, this indicates the passive error of the nasal realisation of /s/, however if nose-holding results in the inhibition of all airflow, this is indicative of an active nasal fricative (Sell et al. 1999).

Other passive errors include absent pressure consonants and gliding of fricatives or affricates. As this thesis focuses on the use of feedback tools for lingual articulations, passive errors, and any assessment or treatment methods for these errors, will not be further discussed.

While the literature suggests that the compensatory error patterns found in speakers with CP are adopted to facilitate phonological development, it could be argued that compensatory articulations are also a result of incorrect motor plans if they persist post-surgery. Preston et al. (2014) suggest that inappropriate phonetic realisations occur due to an inappropriate motor plan. While referring to primary Speech Sound Disorders (SSDs), Cleland et al. (2015c) suggest that erroneous motor plans can be subcategorised into three groups: 1) realisations that are identical to another phoneme, resulting in homophony (e.g. in the case of velar fronting); 2) when the motor plan is abnormal or underspecified, which results in a realisation that sound homophonous but differs in some way (e.g. covert contrast, Gibbon and Scobbie 1997); and 3) when the motor plan is abnormal in the way that results in the realisation of an obviously non-native sound. In the case of children with CP, compensatory errors that persist post-surgery could be assigned to each of these three categories. Errors such as retraction of alveolar to velar placement could essentially result in category 1, with both /t/ and /k/ resulting in homophony and realised as [k]. Errors such as double articulations could result in category 2, where perceptually there may be homophony but there are covert errors such as double articulation. Errors such as palatalisation, pharyngealisation, and glottalisation would be categorised as category 3 (non-native errors). It is crucial, diagnostically, to identify the nature of impairment, whether it be phonological or motoric, to determine the correct treatment plan. The following sections address the assessment and treatment of speech in children with CP, including the principles of motor learning.

1.3 Assessment of Speech in Individuals with Cleft Palate

Assessment of a child with CP will begin in early infancy and will be continuous until adulthood (Scherer and Louw 2011). Due to the heterogeneity of the communication skills in individuals with CP, and with evidence of CTRs appearing at around six months of age (Harding and Grunwell 1996), a broad framework for assessment must be employed (Scherer and Louw 2011). Assessment should include a detailed case history and hearing assessment due to the frequency of conductive hearing loss and otitis media in individuals with CLP (Stengelhofen 1989). Due to the structural nature of CLP, an orofacial examination is also crucial. As children with CP are also at risk of language delay, a language screen must be implemented at an early stage and monitored closely throughout their pre-school years (Kummer

2014). For those who have additional risk factors, for example a conductive hearing loss, a comprehensive language assessment should be carried out (Kummer 2014).

More importantly, a speech assessment must be implemented in children with CP, whether this is perceptual, instrumental, or a combination of both.

In the UK, the most well recognised assessment for the speech of individuals with CP is the Great Ormond Street Speech Assessment (*GOS.SP.ASS*; Sell et al. 1994; revised 1999), which is based on the outcomes from Harding and Grunwell's (1996) reconsiderations of the speech characteristics of individuals with CP, in the context of phonological development (Harding and Grunwell 1998). GOS.SP.ASS'98 was selected by a panel of experienced SLTs as the preferred protocol for the assessment of speech associated with individuals with CP in the UK (Harding and Grunwell 1998; Sell et al. 1999). It assesses both the articulatory effects of the cleft and the idiosyncratic phonological processes commonly associated with CP, for example backing. It demonstrates the importance of including a phonological analysis into routine assessment of the speech of individuals with CP in order to demonstrate phonological contrasts and the extent to which they are affected by the impaired structure and function of the palate (Harding-Bell and Howard 2011). However, it does only provide a partial phonological analysis.

When transcribing speech in those with cleft palate, it is important to consider not only the cleft type characteristics, i.e. the compensatory and obligatory errors (see below) due to the structural abnormalities in the vocal tract, but also the child's phonological system. Previous studies investigating phonological development in speakers with CP have identified errors, or patterns, that are both directly related to the effects of CP and/or to typical development of phonology (Hodson et al. 1983; Lynch et al. 1983). It is suggested that those processes which are related to typical development, for example velar fronting, tend to persist longer in children with a CP than those without CP (Chapman and Hardin 1992; Chapman 1993). As a result, assessment protocols for the speech of individuals with CP should incorporate phonological measures (Morris and Ozanne 2003). Other standardised speech assessments, such as the Diagnostic Evaluation of Articulation and Phonology (*DEAP*; Dodd et al. 2002) provide a more extensive evaluation of a child's phonological system. Using the DEAP phonology subtest for assessment of speech in individuals with CP can be useful in differentiating the cleft-type characteristics from developmental phonological processes.

Kuehn and Moller (2000) suggest perceptual assessment has the greatest face validity for the assessment of speech in individuals with CP and it is considered a key outcome measure in CLP management (Lohmander and Olsson, 2004; Sell, 2005). In most cases, analysis of narrow phonetic transcription is deemed to be gold standard in diagnosing SSDs associated with cleft palate (Sell 2005; Peterson-Falzone et al. 2006; Howard 2011). However, phonetic transcription is subjective and not always reliable. While we try to make our transcriptions as narrow and detailed as possible, this results in inter-transcriber disagreement (Howard 2011). Previous literature suggests that reliability for broad phonetic transcription is most likely to be within the 90–95% range and for narrow transcription is usually around 80% (Shriberg and Lof 1991; Shriberg et al. 1997). However, speech in individuals with CP is particularly complex, and has been shown to have poorer listener agreement than Shriberg et al.'s levels, with lower levels of agreement at only 19% and even the highest level of percentage agreement being below the 80% level at 71%, with an average of 40% (Gooch et al. 2001). As transcriptions from single transcribers can be unreliable, previous reviews of the literature have suggested that multi-listener judgements are the preferred method of choice in the evaluation of speech outcomes in CP speakers (Kuehn and Moller 2000; Lohmander and Olsson 2004; Britton et al. 2014). Various methods of multi-listener perceptual evaluations have been adopted for measuring the speech of individuals with CP, with the majority of perceptual evaluations investigating resonance disorders (Prathanee et al. 2012; Baylis et al. 2015). Both Ordinal Scales, such as those found in the CAPS-A (Sell et al. 2009), and visual analogue scales (*VAS*; Munson et al. 2012; Baylis et al. 2015) have been found to be effective tools for measuring speech outcomes in CP (Castick et al. 2017). Section 3.1 provides further detail on perceptual evaluations for individuals with CP.

By using a range of formal and informal assessments to investigate the speech of individuals with CP, differential diagnosis can be made and management plans can be tailored to the specific articulation errors or phonological processes of each child.

The following section will discuss approaches used for therapy for speech disorders associated with CP.

1.4 Treatment of Speech in Individuals with Cleft Palate

Active/compensatory errors employed by speakers with CP have been previously discussed, along with the suggestion from Harding and Grunwell (1998) that these active errors require SLT input. SLTs will usually encounter the child after initial surgical treatment, however treatment may continue into teenage years. Therefore, it is necessary for SLTs to be aware of any planned surgical intervention which may affect the treatment plan (Peterson-Falzone et al. 2006).

Children with CP will typically undergo therapy in their pre-school years. However, some children will begin therapy earlier or later in their school years, for example if symptoms of VPD are not evident until further palatal repair. Extended periods of intervention are likely for individuals with CP. Intensive therapy, in both individual and group contexts, has shown to be effective (RCSLT 2005). Peterson-Falzone et al. (2006) suggest that a realistic therapy schedule should consist of twice weekly sessions, lasting approximately 30 minutes per session, supplemented by daily homework. It is appropriate to have the goal that individuals with CLP should achieve age appropriate articulation, voice quality and adequate language skills (Van Denmark 2004). Kummer (2014) supports this and notes that the goal of speech therapy is to correct placement, and occasionally manner, of articulation.

Due to the articulatory nature of CP, articulation therapy is the preferred approach for individuals with CP (Peterson-Falzone et al. 2010). Traditional articulation therapy (Van Riper 1978) is well known for its hierarchy of activities (Van Riper and Emerick 1984). It is essentially a motor-based approach, focussing on articulatory movements. Two key components within this intervention are self-monitoring and the ability to self-correct (Peterson-Falzone et al. 2006). The process begins with sensory-perceptual training, which focuses on identifying a sound and discrimination tasks. It then moves on to varying and correcting various productions of the sound until produced correctly. The third phase is strengthening and stabilising the correct production, and finally, the fourth phase focuses on transferring into spontaneous, everyday communications. There are also four successive levels in which the process is implemented:

1. The isolated sound level
2. The sound in a syllable, for example consonant vowel (CV), VC, VCV, CVC
3. The sound in a word
4. The sound in a meaningful sentence (Van Riper and Emerick 1984).

With the argument that compensatory errors post-surgery are essentially incorrect motor plans (Preston et al. 2014; Cleland, et al. 2015c), this would suggest the requirement of motor-based intervention, such as traditional approaches (Van Riper and Emerick 1984; Van Denmark and Hardin 1986; Pamplona et al. 2005) or visual biofeedback approaches such as Electropalatography (EPG; Lee et al. 2009) or ultrasound (see section 1.5). The following sub-section addressed the theory of motor learning and motor-based intervention.

1.4.1 Motor Learning

Maas et al. (2008) suggests that the principles of motor learning in non-speech motor control (non-speech oro-motor movements and other activities such as golf) are similar to those motor skills necessary for speech production, and that the structure of practice, target selection and the nature of the feedback provided can facilitate the acquisition of a new sound. The key aim in motor-based therapy is to ensure that the acquisition (or performance) of a new motor skill is retained and generalised over time.

In motor learning, two phases are required: pre-practice and practice. Each therapy session begins with a period of pre-practice, where the child is taught how to produce the target phone. It is during this pre-practice phase that acquisition occurs. To achieve *acquisition* of a speech sound, the trained gestures should be practiced and shaped in to new movements. For example, in the case of acquiring a velar stop in a range of vowel environments a facilitative environment [ok] may be used to shape a new gesture to acquire a velar stop in a number of vowel environments. Once a sound is acquired, the new skilled movement should not only be observed during practiced items but retained over time. The term *retention* refers to performance levels once practice is completed and *generalisation* is used when the new gesture is transferred to untrained movements that are related (Maas et al. 2008). Generalisation can be measured by including an untreated wordlist into speech materials. Both retention and generalisation occur during the practice phase of intervention. The conditions of practice influence how well the new motor skill (speech sound) is acquired, retained and generalised to other sounds or contexts.

1.4.1.1 Practice Conditions

Practice Amount (Dosage) refers to the amount of practice provided during therapy sessions (for example small vs. large amounts of practice). Schmidt and Lee (2005) suggest that the more practice occurs, the more learning occurs, therefore suggesting that for motor-based speech therapy, large amounts of practice (high dose) should be provided during every session.

Practice Distribution refers to how the amount of therapy is scheduled or dispersed over time (Maas et al. 2008). Practice can be massed, i.e. with less time between trials and/or sessions, or distributed, i.e. more time between trials and/or sessions. It is suggested that massed practice is useful for initial acquisition of a new skill and that distributed practice is more helpful for retention and generalisation (Caruso and Strand 1999; McLeod and Baker 2017).

Practice Variability refers to whether practice is constant or variable, i.e., whether the same target is practiced in the same context or whether a target is practiced in different contexts. This can be linguistic contexts, such as word positions or vowel environments (Skelton and Hagopian 2014) or the same skill within variable parameters such as change in rate, pitch, force or intensity (Preston et al. 2014). Preston et al. (2014) suggest that constant practice is useful when initially acquiring a new motor skill (i.e. speech sound). However, in order to achieve retention and generalisation, i.e. true learning, therapy should shift toward using variable practice.

Practice Schedule can be either blocked or random, i.e. practicing one speech target a number of times prior to moving onto the next target vs. presenting the targets in random order to prevent the learner from predicting what trial follows another (Maas and Farinella 2012). It is suggested that a mix of both blocked and randomised practice may be useful for learning a new speech sound (McLeod and Baker 2017).

Practice Complexity refers to whether a task is simple (e.g. the target in isolation) or complex (e.g. the target in a complex cluster or within a sentence. Similar to that of the hierarchy presented by Van Riper and Emerick (1984), motor-based therapy should increase the level of complexity of tasks over the course of a block of therapy. With the view that motor tasks represent a variety of challenges for performers of various abilities, the Challenge Point Framework, proposed by Guadagnoli and Lee (2004), suggests that increases in the level of difficulty in tasks may increase learning potential, however if tasks are increased past the skill level of the learner, their performance will decrease. Therefore, the optimal 'challenge point' occurs when learning opportunities are maximised but the detriment to performance practice is minimised.

Recent studies using ultrasound visual biofeedback (Preston et al. 2014; Cleland et al. 2015c; Cleland et al. 2017c) use a therapy hierarchy which fits with the Challenge Point framework with an increasing level of complexity in tasks for acquisition (elicitation), retention and generalisation, with an 80% step-up criteria. Table 3 shows the therapy hierarchy used in Cleland et al. (2017c), ranging from CV or VC using a facilitative vowel, through to a range of CV or VC syllables (level 2) to CVC monosyllabic (level 2) and polysyllabic words (level 3). Levels 4 and 5 focus on singleton tokens in phrases and sentences, while level 6 targets complex clusters. Finally, level 7 focuses on practicing the target sound in complex sentences for generalisation.

Level 0	CV or VC facilitative vowel
Level 1	CV
Level 1	VC
Level 2	CVC WI
Level 2	CVC WF
Level 3	Multisyllables
Level 4	Phrase repetition WI
Level 4	Phrase repetition WF
Level 5	Cloze (sentence completion)
Level 6	Clusters
Level 7	Complex sentences repetition and invention

Table 3 Hierarchy for Therapy Levels with 80% step-up criteria (used with permission, Cleland et al. *2017c)*

Practice Fraction refers to whether a motor skill is practiced as a whole movement or parts of a movement. Skelton (2004b) suggest that part practice of some speech movements is possible. There are three subtypes of part practice: simplification, segmentation and fractionation (Wightman and Lintern 1985). Simplification (i.e. making a skill easier) during pre-practice may aid acquisition of the whole motor skill. Segmentation involves breaking down a skill along a temporal dimension, for example by breaking a word down into syllables or phonemes, and then blending them together. Fractionation involves breaking down the simultaneous movements (e.g. lingual gestures) and practicing each movement (Forrest 2002), which results in non-speech oro-motor movements (i.e. elevation of the tongue tip to touch the alveolar ridge without making a [t] sound). McLeod and Baker (2017) suggest using segmentation and simplification rather than fractionation, as non-speech oro-motor exercises are not recommended for children with SSD (Forrest 2002). However, fractionation should not be confused with shaping a new articulatory gesture, i.e. providing gestural instruction to elicit or acquire a new speech sound. When using biofeedback techniques such as Electropalatography or ultrasound, learners are able to watch a video or image of a target tongue shape and manipulate their tongue shape in real-time to achieve a target articulatory gesture. For some children, they may not be able to do this from just looking at a target and may require shaping, in the form of the SLT providing gestural instruction along with the biofeedback (e.g. by instructing the learner to move specific parts of the tongue to touch areas on the ultrasound screen).

Practice Accuracy refers to errorless and errorful learning. Within errorless practice, errors or mistakes are discouraged, which in turn supports accurate acquisition or a new motor skill. However, for children with CP this may not always be possible due to any anatomical abnormalities. Therefore, a client-specific variant of errorless practice may be required to ensure that errors are minimised rather than removed completely as this may be structurally impossible in the case of obligatory errors. Errorful practice provides opportunity for mistakes, in turn providing a chance for the learner to identify their own errors and self-correct (Maas et al. 2008; Bergan 2010). Motor-based interventions should allow for mistakes and provide opportunities for learners to refine their articulations and ability to self-correct (McLeod and Baker 2017).

Attentional Focus during pre-practice and/or practice can be external or internal. Instructions focusing on the acoustics of an articulatory movement are said to be external, while those that focus on the articulatory gestures (i.e. instructions that focus the child's attention to how the sound is made) are said to be internal (Maas et al. 2008). During pre-practice, it is likely that the SLT will use both internal and external instructions or feedback, whereas during the practice phase, it is more useful to use external focus (Lisman and Sadagopan 2013; McLeod and Baker 2017). Recent studies (Rvachew and Brosseau-Lapre 2012; Hitchcock and McAllister Byun 2015) suggest that using biofeedback to manipulate conditions of practice could allow for greater gains in generalisation. This will be discussed in more detail in section 1.5.

1.4.1.2 Feedback Conditions

As well as the conditions of practice within the motor learning principles, a distinction is also made between the types of feedback, the feedback frequency and the timing of feedback used during therapy.

Two types of feedback are considered in motor learning: Knowledge of performance (KP) and knowledge of results (KR) (Preston et al. 2014). KR refers to feedback related to the results produced in terms of target correctness and would be provided after a movement is completed. KR feedback can be provided by the SLT, through auditory feedback such as “well-done that was a really good [k]”. In contrast, KP refers to feedback on *how* the gesture is produced. While this can be provided by an SLT, for example, “the front of your tongue was up”, this type of feedback cannot be accurately provided for gestures that are not viewable for the treating clinician (such as velars), therefore biofeedback tools, such as ultrasound (Preston et al. 2014; Cleland et al. 2015c; Cleland et al. 2017c) have been used to provide this feedback directly to the learner and also for the SLT to provide more accurate verbal feedback. Both KP and KR feedback are useful in the pre-practice phase of motor-based therapy. During the practice phase, more KR feedback should be provided.

Feedback frequency refers to how often feedback is provided (Maas et al. 2012) and can be either high- or low-frequency. High-frequency feedback refers to providing KP or KR feedback after every trial, whereas low-frequency is when feedback is provided on 50% or fewer trials (McLeod and Baker 2017). Maas et al. (2008) suggest that high-frequency feedback is useful in the pre-practice phase, as it enhances performance. Low-frequency encourages the learner to use their own intrinsic feedback and is said to be more useful in the practice phase of intervention (McLeod and Baker 2017).

As well as feedback frequency, feedback timing is also considered. Maas et al. (2008) suggest that feedback can be concurrent with a response, immediately after the learner’s response or following a short delay. Both concurrent and immediate feedback will help improve performance during practice; however, delayed feedback is believed to be more helpful for learning as the delay encourages learners to detect any errors and self-correct, therefore improving on their next attempt (Ballard et al. 2010; Murray et al. 2015; McLeod and Baker 2017). Using a bespoke therapy version of Articulate Assistant Advance (AAA; Articulate Instruments 2012) ultrasound software, recordings can be made of a child’s attempts of a target. During the recording, no KP or KR feedback would be provided but recordings would be played back to the child immediately afterward. This type of delayed feedback (not to be confused with delayed auditory feedback), allows the child to use their internal feedback system to watch and rate their own productions, in turn self-correcting and improving their next attempt. It also allows the child and SLT to discuss any errors together.

1.4.2 Efficacy of Motor Based Therapies for Individuals with Cleft Palate

Bessell et al. (2013) carried out a systematic review in order to examine the evidence of differences in timing and type of SLT and to identify the types of interventions implemented. Their selection criteria included both randomised and non-randomised control trials, patients with CP+/-L (syndromes were included if there were no known developmental delays), any SLT interventions and any speech outcome. A no intervention control group of a different SLT intervention were used for comparisons. Out of 17 papers, only six were randomised trials, with the remaining 11 being observational studies. With regards to intervention types, 10 evaluated motor approaches and seven evaluated linguistic approaches, for example, phonological approaches such as minimal pairs. Based on the data reported, Bessell et al. (2013) concluded that it was difficult to determine which approach was more effective, with little evidence to support any one intervention approach (motor approaches such as articulation therapy, or linguistic approaches such as phonological, whole language or focus stimulation approaches), in relation to duration, setting, intensity, delivery, age of intervention or theoretical perspective. Similarly, Meinusch and Neumann (2016) highlight the need to compare motor approaches to linguistic approaches. Vallino-Napoli (2011) also notes that the literature is insufficient in providing evidence for the direct effects of various intervention approaches. It was also highlighted in this systematic review that there is a requirement to investigate the impact of therapy on a child's communication and psychosocial wellbeing.

Van Denmark and Hardin (1986) tested the effectiveness of a six-week residential programme of articulation therapy for 13 children, aged six to 12, with CLP or CP. Within this intensive programme, they implemented the systematic multiple-sound approach to articulation therapy described by McCabe and Bradley (1975). Within this approach, multiple sounds are introduced within each session plan, despite the child's level of performance on each sound (Van Denmark and Hardin 1986). Each child was provided with four one-hour sessions of therapy for 26 days, providing 104 hours of therapy in total. Results showed significant improvement in articulation scores, however slow progress was noted. A nine month follow up assessment showed no significant gains during the maintenance phase across participants, with only three participants showing improved skills in the maintenance assessment (Van Denmark and Hardin 1986).

Pamplona et al. (2005) also report on an intensive summer camp for treating articulation disorders in children with CP. Their study adopts a whole-language model, in which phonological principles are incorporated. Two matched groups, each with 45 children with repaired CP and compensatory articulation disorder, were included. Using the same model of intervention with particular focus on reducing glottal and pharyngeal articulations, the first group received two one-hour sessions of therapy per week for 12 months, whereas the other group received four hours of therapy per day, five days a week for three weeks. Results show that both groups had a significant decrease in their severity of compensatory articulations.

The two studies described above provide moderate evidence to support RCSLT's recommendation for intensive therapy for working on articulation. Pamplona, et al. (2005) showed no significant differences between an intensive programme and a more conventional approach of two sessions a week, also suggested by Peterson-Falzone et al. (2006). More realistically, in a clinical setting, it is more likely to be one session per week, maximum, due to time and financial constraints. NHS services are mostly based on the impact of the child's difficulties on intelligibility or well-being and the clinical need for SLT intervention. Each health board have their own pathways and policies for service provision, therefore there is no consensus regarding time allocation per client.

Van Denmark and Hardin's approach to therapy is based on an articulatory approach, whereas Pamplona et al.'s approach incorporates phonological principles into a whole-language approach. This highlights the importance of phonological approaches as well as articulation approaches, and relates back to Harding and Grunwell's (1996; 1998) suggestion of reconsidering the speech of individuals with CP in a phonological context. Using a phonological approach will also address any errors found in the phonological analysis incorporated in GOS.SP.ASS'98 (Sell et al.

1999) or other standardised assessments such as the DEAP (Dodd et al. 2002). Phonological intervention aims to expand the child's speech sound system, along with the syllabic and lexical contexts of which target sounds may occur, in turn improving intelligibility (Harding-Bell and Howard 2011).

Using a prospective, randomised trial, Pamplona et al. (1999) investigated whether using a phonological approach would reduce the total time of speech therapy treating compensatory articulations, compared to an articulation approach. One group of children (N=14) received phonological intervention and the other group (N=15) received articulation therapy. All participants received two one-hour sessions weekly until they had achieved articulation within normal limits. Results showed that speech therapy time was significantly reduced in the group receiving phonological intervention (mean total time in therapy = 14.5 months) compared to those receiving articulation therapy (mean total time in therapy = 30.07 months).

Up until now, an overview of assessment and therapy approaches for speech disorders in individuals with CP has been provided, with regards to articulation and phonology. In assessment of speech for individuals with CP, phonetic transcription is deemed to be gold standard (Sell 2005; Peterson-Falzone et al. 2006; Howard 2011), with transcription considered a key outcome measure of treatment (Lohmander and Olsson 2004; Sell 2005). However, transcription does have its disadvantages. It can be subjective and not always reliable, with more detailed or narrow transcriptions leading to inter-transcriber disagreement (Howard 2011). Due to the subjectivity of phonetic transcription, it has been suggested that instrumental articulatory tools, such as EPG and UTI, have shown promise as an aid for assessment of articulatory movements in the speech of individuals with CP. Although it is reported that using conventional approaches to therapy has shown improvements in speech outcomes, for some children, using these conventional approaches have been unsuccessful. As a result of this, the same instrumental tools have also been used for visual biofeedback as an adjunct to therapy as well as for assessment (Sell 2005; Peterson-Falzone et al., 2006; Howard 2011). The remainder of this chapter will discuss the visual feedback technologies used to image the articulators and their application for individuals with CLP.

1.5 Visual Articulatory Feedback Technologies

Visual articulatory feedback systems aim to provide speakers with visual information regarding their own articulation (Youssef et al. 2011). A range of visual feedback technologies (see Table 4) have shown promise as useful tools for assessment and therapy for children with speech sound disorders, including those associated with CLP.

Auditory/ Acoustic Feedback Techniques	Acoustic Biofeedback Techniques	Tactile Feedback Techniques	Visual Feedback Techniques	Visual Biofeedback Techniques
SLT commenting on KR (e.g. "well done, that sounded like a really good [k]") Delayed auditory feedback (for example in the case of stammering Van Borsel et al. 2013) Amplification of target sounds (e.g. Hodson 2010) Listening to a recording of their own speech	Spectral Feedback (e.g. McAllister Byun and Hitchcock 2012) See Scape (Pro-Ed 1986) Nasometry (Fletcher 1972) Perci-sars (MicroTronics Corp 2016) LingWaves TheraVox (Wevosys 2017) Nasopharyngoscopy (Neumann and Romonath 2012)	To provide feedback on correct tongue placement and lingual coordination (Altshuler 1961; Shriberg 1980), e.g. Tongue depressors Peanut butter Lollypops	Schematic illustrations for teaching tongue placement (e.g. phonetic textbook illustrations) Cued articulation for placement and manner (Passy 2010) Visual Articulatory Models/Talking Heads (e.g. Kroger et al. 2005; Badin 2010; Lawson et al. 2015)	Electropalatography (Lee et al. 2009) Mirror Electromagetic Articulography (Katz et al. 2010) Ultrasound (see Table 5)

Table 4 List of the techniques used by SLTs to provide feedback to clients during therapy

While acoustic biofeedback techniques, such as Perci-Sars (MicroTronics Corp 2016) and Nasometry (Fletcher 1972), are widely used for the remediation of resonance disorders in children with CP, the focus of this thesis is articulatory disorders and not resonance disorders. Therefore, the remainder of this section will focus on visual biofeedback techniques only.

It is important to distinguish visual feedback, i.e. feedback or instruction concerning visual information, such as a drawing or the use of a Visual Articulatory Model (VAM), from visual *bio*feedback. France and DeAngelo (2016) describe biofeedback, in its most general term, as an individual's ability to learn how to self-regulate and change automatic physiological functions that they are not consciously aware of through a monitoring device. While speakers are consciously aware of the articulators, these are largely hidden, with only the face, lips, teeth and tongue tip visible (Cleland et al. 2013), which leads to difficulties describing or modelling the articulatory gestures when learning a new sound, an essential aspect of motor-based therapy. Visual biofeedback techniques, such as electropalatography (EPG) and ultrasound visual biofeedback (UVBF) can circumvent these difficulties by augmenting the acoustic and tactile feedback used in more traditional therapeutic methods, by allowing the learner to view their own tongue moving in real-time, to self-regulate and change their tongue-shape to learn a new motor plan.

For cleft-type errors, arguably impaired motor-plans such as retraction or double articulations, instrumental articulatory analysis may be essential for diagnostic purposes (i.e. to detect specific lingual errors) and biofeedback used for therapeutic purposes (for speakers to monitor and self-regulate articulatory gestures in real-time and for SLTs to provide more accurate feedback during therapy). Although visual biofeedback techniques such as EPG are well established, with a vast evidence base in the cleft palate literature, UVBF is growing increasingly popular, with improved methods and lower costs (Stone 2010). As EPG is well established within the literature for treating SSDs in children with CP, the following sections address the use of EPG, and newer techniques such as UVBF and VAMs, for assessing and treating SSDs associated with CP and their role in motor-learning.

1.5.1 Electropalatography

Electropalatography (EPG) measures tongue-palate contact in real-time during speech and speakers use this visual information to modify erroneous articulations. Speakers wear a custom made artificial palate (Figure 3) which is approximately 1-3mm thick and contains 62 electrodes. Row one of the palate contains six electrodes which are situated closely together next to the upper incisors and rows two to eight contain eight electrodes which are more widely spread at the junction of the hard and soft palate. When the tongue makes contact with an electrode on the palate, a signal is sent to the processor and the pattern of contacts is schematically displayed on a computer screen.

Figure 3 Reading EPG Palate

Each row of electrodes is represented as a row of squares on the computer screen, with a filled square representing tongue-palate contact at that point (Figure 4) (Stone 1999; Hewlett and Beck 2006). The same schematic pattern is used for all shapes of palates, even those which, due to clefting, have a distorted or asymmetrical shape (Figure 4). In other words, all speakers see the same display irrespective of the size or shape of their palate. For some speakers with CP, not all electrodes can be placed on the palate due to a high arched or narrow palate. While EPG does not directly show the direct result of the cleft, it is able to provide a comparison of the data from speakers with CLP to normative data due to the normalisation of patterns in the examples below. This makes the display easier for speakers and SLTs to interpret, in turn making it easier for SLTs to provide more accurate gestural instruction and feedback.

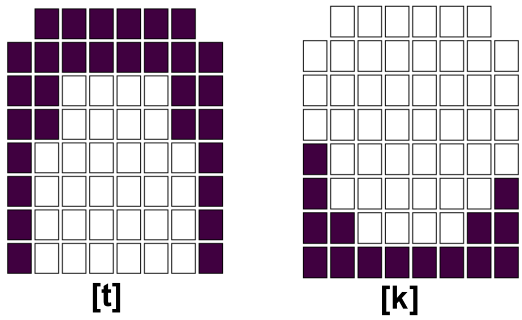


Figure 4 Tongue-Palate Contact represented on computer screen (same normalised representation for typical and CP anatomy)

As EPG requires individualised palates, this makes it expensive to implement as each palate currently costs around £500 to make and can only be used when a client is in a stable period of dentition. Therefore, there is a limited period in which therapy can be implemented using EPG, particularly for individuals with CP as they are likely to require additional orthodontic treatment or secondary surgery. They also have unpredictable dental eruption which sometimes prevents the EPG palate being used. This is not the case with Ultrasound Tongue Imaging (UTI), which can be used regardless of dentition. Another disadvantage of EPG is that the plate only extends as far back as the velum. As many of the reported speech characteristics in CP are retracted further than the velum, these would show up on an EPG screen as an open pattern (as described by Gibbon 2004), therefore no information on lingual movement would be visible. The following section discusses the error patterns identifiable using EPG and the evidence based for EPG treatment for speakers with CP.

1.5.1.1 Electropalatography and Cleft Palate

Previous studies have used EPG to assess speech in individuals with CP (Gibbon 2004; Gibbon et al. 2004) and have investigated the therapeutic applications of EPG in children with CP (Lee et al. 2007). A Cochrane Review of EPG for articulation disorders associated with CP (Lee et al. 2009) summarised that, despite recommendations from RCSLT (2005), there is a lack of evidence to support the efficacy of EPG in treating articulation disorders in CP. They recommend that randomised control trials be undertaken before this technique is used routinely for individuals with CP (Lee et al. 2009).

In the speech of individuals with CP, active compensatory articulations occur due to VP causing difficulty producing high pressure consonants (Harding and Grunwell, 1998), in turn leading to impaired motor-plans. These compensatory articulations are often characterised by posterior placements not normally found in English, for example pharyngeal or glottal stops (Trost, 1981; Harding and Grunwell, 1998). While glottal stops are commonly found (mostly in word medial or final position) as the phonetic realisation of /t/ in many variations of English (Smith and Holmes-Elliott 2017) in children with CP, they may also retract word initial /t/ to glottal placement and also retract velar stops to glottal placement, which is not typical within English. These types of errors are not imageable with EPG (since it samples only as far back as the juncture of the hard and soft palate) and are displayed only as an open pattern, however the target consonant (at least in English) is imageable, making it possible to use EPG for identifying some of the CTRs outlined in sub-section 1.2.3 and for biofeedback. Secondly, EPG is not suitable for all clients with CP due to requirements for secondary surgery or on-going dental, orthodontic or maxillary input. The following sub-section will discuss the compensatory errors identified through analysis of EPG data, followed by a discussion of the evidence of therapeutic studies using EPG.

1.5.1.1.1 EPG for the Assessment of Speech in Individuals with Cleft Palate

Howard (2004) and Gibbon (2004) have highlighted the importance of instrumental analysis in the assessment of speech associated with CP. In her 2004 study, Howard reported covert errors, such as dorsalpalatal contact, retraction, double articulations and increased intra- and inter- speaker variability. Similarly, Gibbon (2004) summarises the abnormal patterns of tongue-palate contact in the speech of individuals with CP, which have been identified in 23 published articles. Firstly, Gibbon (2004) identified *increased contact*, which could affect all lingual consonants and vowels. Although it remains uncertain why there is increased contact in speakers with CP, Hardcastle et al. (1989) proposed that speakers with CP have an impaired development of typical tongue function due to palatal scarring and therefore, have a lack of tactile awareness. Other possible causes could be due to the presence of fistulae, hearing impairment, concomitant verbal dyspraxia or compensatory actions of the tongue apex (Hardcastle et al. 1989). Previous studies of CP speech using EPG (Morley 1970; Lawrence and Philps 1975; Golding-Kushner 1995) have also identified overuse of the tongue dorsum. A possible reason for the overuse of tongue dorsum is to aid velopharyngeal function (Trost, 1981). Overuse of the tongue dorsum is also associated with the second pattern identified by Gibbon (2004): *retraction to palatal or velar placement*, or mid-dorsum palatal stops (Trost, 1981). Palatalisation of fricatives is previously reported in the EPG literature (Michi et al. 1990; Yamashita and Michi 1991; Howard 1998; Howard and Pickstone 1995). It is suggested that increased contact, or overuse of the tongue dorsum, can result in reduced control of the lateral margins of the tongue (Gibbon 2004). Control of the lateral margins of the tongue is essential for speech production (Stone et al. 1992). Gibbon (2004) proposes that if there is increased contact, this will in turn have a significant and detrimental effect on the development of speech motor control.

Another compensatory articulation commonly reported in the literature is retraction to uvular or pharyngeal placement. EPG alone is unable to identify these errors as the EPG palate only goes as far back as the velum. Gibbon (2004) describes this as an *open pattern* where there is no tongue-palate contact identified on the EPG plate but there may be in more posterior regions, for example, uvular, pharyngeal or glottal placement. Gibbon suggests that these open patterns, or retracted articulations, are likely to be caused by the motor control of the tongue tip/blade and tongue body developing in an atypical fashion, due to the lack of experience of typical tongue movements. *Fronted placement,* another atypical pattern identified, typically involves contact in the palatal region for velar consonants. This error is less common in speakers with CP than the process of retraction, with Spriesterbach et al. (1956) suggesting that velars are less likely to involve placement errors than alveolar targets. Alongside fronting and backing, Gibbon (2004) discusses *reduced separation of alveolar and velar placement*. A previous study by Gibbon and Crampin (2001) found that although there was a reduced separation between alveolar and velar targets, where both targets were perceptually neutralised and transcribed as the mid-dorsum stop [c], a statistically significant contrast was identified in the articulatory data. Gibbon and Crampin highlight the importance of distinguishing between reduced separation and contrast neutralisation for diagnostic purposes as this may differentiate errors which are phonetic in nature to those that are phonological in nature. Similar to Howard (2004), Gibbon also reports *double articulations* as an atypical error in speakers with CP. However, as the palate only goes as far back as the velum, only alveolar/velar double articulations or alveolar/glottal or velar/glottal articulations will be visible. There may also be *abnormal timing* in the control of articulators which can be identified through temporal measurements of EPG.

The penultimate pattern Gibbon identifies is *complete closure,* when the tongue comes into contact with all 62 electrodes on the palate. In the literature on primary SSDs, Gibbon (1999) also describes undifferentiated lingual gestures (ULG) as having a high amount of closure, with lack of a clear distinction between the tongue apex, tongue body and lateral margins on the tongue. She suggests that ULGs are a result of speech motor constraint with delayed or deviant control of the independent regions of the tongue.

Similar to Howard's finding of increased intra-speaker variability, *increased variability* was also identified as an atypical pattern in Gibbon's 2004 paper as the final error. This is described as variability across repetitions of the same word in location and amount of contact over different regions of the EPG palate. Patterns for the target sound may also vary in different vowel and consonant environments (Hardcastle et al. 1989).

It is evident that EPG is useful for detecting covert lingual errors for diagnostic purposes. The RCSLT (2005) also recommends using EPG for treating SSDs in individuals with CP.

1.5.1.1.2 EPG for the Treatment of Speech in Individuals with Cleft Palate

Typically, EPG is used in conjunction with an articulation therapy, or motor-based, therapy approach (Gibbon et al. 2001; Lohmander et al. 2010). Lee et al. (2009) carried out a Cochrane Review of the use of EPG for treating articulation disorders associated with CP. The search identified 112 studies. Using the criteria of a randomised control trial comparing EPG to no treatment, delayed treatment, conventional therapy, or alternative treatments, only one study met their criteria

(Michi et al. 1993), which included six Japanese speakers with compensatory articulations. Using a parallel group design, Michi et al. (1993) compared three treatments (EPG and friction display method, EPG therapy and conventional articulation therapy). While no statistical analysis was carried out, Michi et al. reported that fewest therapy sessions were required in the EPG and friction display method condition, with conventional therapy requiring the highest number of therapy sessions.

Although not included in Lee et al.'s Cochrane Review, two studies, CLEFTNET Scotland (Gibbon et al. 1998) and the national CLEFTNET project (Lee et al. 2007) are well known for linking CP centres within the UK to provide individuals with CP access to EPG therapy. Gibbon et al. (1998) conclude that CLEFTNET Scotland demonstrated the value of an electronic network for facilitating the use of EPG for individuals with CP, by providing technical support for clinicians, providing a detailed data analysis, and allowing clinicians access to expertise in the use of EPG. They found that EPG therapy was effective for identifying lingual errors and also improving compensatory articulations. The national CLEFTNET project (Lee et al. 2007) widened this service to UK-wide cleft centres, finding similar results to those in CLEFTNET Scotland. From this study came the development of therapy guidelines for clinicians, which followed a motor-based approach and the use of facilitative contexts (Kent 1982) for learning new articulations. However, Lee et al. (2007) identified problems with EPG therapy, due to specific dental problems associated with CP, including the need for a period of stable dentition which is problematic in speakers with CP due to the late development of their secondary dentition. They also reported timing of surgical intervention or ongoing orthodontic treatment as problematic, as it could interfere with EPG therapy. There is also the possibility that EPG palates can break and are expensive to have repaired or replaced. Another such tool that may overcome some of the issues with EPG, is Ultrasound Tongue Imaging (UTI).

1.5.2 Ultrasound Tongue Imaging

Ultrasound Tongue Imaging (UTI) works by using reflective properties of sound waves, which travel through soft tissue. When the sound waves reach a tissue of a different density (e.g. air or bone) the sound waves reflect back creating an image on the screen. In the case of UTI, the ultrasound probe is held under the chin, normally with equally distanced mandible and hyoid shadows. The tongue surface is displayed as a white line on the screen. Typically, UTI uses a 2D image, however 3D systems are available. Two views, midsagittal and coronal, can be used to acquire a tongue image. Using a midsagittal view, we are able to see the tongue's surface and movement patterns from near the tongue tip to the tongue root (Figure 5). However, as UTI does not travel through the jaw or hyoid bones, the tip and root are not always visible.

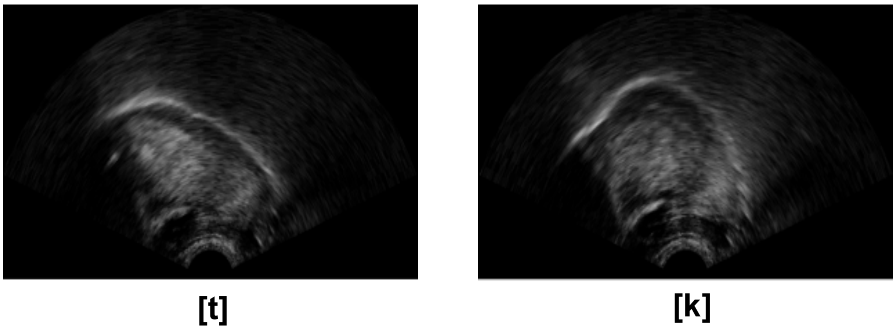


Figure 5 Midsagittal View of 2D Ultrasound Tongue Imaging (tongue tip on the left)

In contrast to EPG, which shows only as far back as the velum, ultrasound images show most of the length of the tongue, from near the tongue tip to the root, with uvular and pharyngeal articulations clearly visible unlike EPG that shows only to velar placement. Furthermore, ultrasound does not require the individualised hardware needed for EPG. Crucial to individuals with CP, there is no requirement for UTI to be used within stable periods of dentition or periods with no planned surgical intervention. Another advantage of ultrasound is that vowels, and therefore the coarticulatory effects of vowels on consonants, are easily viewed.

Similar to EPG, UTI has been, or is currently being, used clinically for both diagnostic and therapeutic purposes, with the use of ultrasound visual biofeedback (UVBF) therapy becoming increasingly popular, with around 30 small number (N, range=1-13) studies published in the field. Table 5 provides a summary of the evidence for the therapeutic use of UVBF. Studies using UVBF for developmental SSDs (DSSD) make up the largest proportion of the evidence (N=14), with /r/

making up over half of therapy targets in the studies (N=16), due to the communicative and social importance of /r/ in North America where most of these studies have been carried out. Twenty-five of the studies only investigated children aged four to 18 years. Three studies included participants up to age 22 with the remaining three studies investigating adults only. Nineteen of the studies showed positive results, with eleven showing mixed results.

Client Group	Number of Studies	Studies	Therapy Targets	Results
Developmental Speech Sound Disorder	14	Shawker and Sonies (1985), Adler-Bock et al. (2007), Modha et al. (2008), Klein et al. (2013), Lipetz and Bernhardt (2013), McAllister Byun et al. (2014) Cleland et al. (2015c), Hitchcock and McAllister Byun (2015), Lee et al. (2015), Bressmann et al. (2016), Heng et al. (2016), Melo et al. (2016), Sjolie et al. (2016)	/r/ (10), velars (2), various (2)	+ve (11), mixed (3)
Hearing Impairment (HI)	5	Bernhardt et al. (2003), Bernhardt et al. (2005), Bacsfalci et al. (2007), Bacsfalvi (2010), Bascfalvi and Bernhardt (2011)	/r/ (1), various (2), vowels (1), N/A (1)	+ve (3), mixed (2)
Childhood Apraxia of Speech	4	Preston et al. (2013), Preston et al. (2016a), Preston et al. (2016b), Preston et al. (2017b)	/r/ (1), various (2), sequencing (1)	mixed (3)
Residual Speech Sound Errors	3	Bernhardt et al. (2008), Preston et al. (2014), Preston et al. (2017a)	/r/ (2), various (1)	+ve (2), mixed (1)
Down’s Syndrome	1	Fawcett et al. (2008)	/r/	+ve
Glossectomy	1	Blyth et al. (2016)	various	+ve
Acquired Apraxia of Speech	1	Preston and Leaman (2014)	/r/	+ve
Cleft Lip and palate	1	Roxburgh et al. (2016)	various	mixed

Table 5 Evidence for the efficacy of UVBF therapy for SSDs (1985-2017) (Adapted from Cleland and Isles 2017 cf Articulate Instruments 2017)

The earliest intervention study of UTI in HI was undertaken in 1985 by Shawker and Sonies. Since then, various studies have implemented the use of ultrasound for HI children and adolescents and have found improvements (Bernhardt et al. 2003; Bernhardt et al. 2005; Bacsfalvi et al.2007; Bacsfalvi 2010; Bacsfalvi and Bernhardt 2011). All of these studies have used either a case study or single-subject design with small participant numbers. Since 2013 there has been a rapid increase in the number of studies investigating the clinical application of UVBF, with 20/30 studies published in the last four years. These studies are mostly Northern American and target /r/, however more studies are now being published investigating a wider range of speech sounds (Lipetz and Bernhardt 2013; Preston et al. 2014; Cleland et al. 2015c; Heng et al. 2016; Melo et al. 2016).

Whilst the number of studies investigating UVBF is increasing, only four studies compare UVBF to other VBF tools, namely EPG. Bernhardt et al. (2003) investigated the use of ultrasound and EPG in four adolescents aged 16-18 with moderate-severe HI. Randomisation to each tool was implemented. Significant improvement for all participants was reported and no advantages were found for either EPG or ultrasound over the other, thus suggesting that UTI is just as effective as EPG and for the CLP may in fact be the preferred tool with its ability to image retracted articulations and no requirement for stable periods of surgical treatment or dentition. As a follow-up to their 2003 study and others (Bacsfalvi et al. 2007; Bacsfalvi 2010), Bacsfalvi and Bernhardt (2011) implemented a perceptual evaluation, whereby the long-term speech outcomes (two-four years post-therapy) were judged by unfamiliar, expert, listeners. Five out of seven speakers maintained their level of performance or continued to generalise post-therapy. Results showed that using EPG and ultrasound as adjuncts to speech therapy provided reduced therapy times, outcomes that were not previously possible with conventional approaches, and provided long-lasting effects on participants' speech outcomes, thus highlighting the need for studies comparing UVBF to other, similar, tools.

Though there is growing evidence for the efficacy of UVBF for SSDs in general, it is clear from Table 5 that there is a lack of evidence for using UVBF with individuals with CP with only one published study, which is in fact the sub-study presented in this thesis in chapter 3. Whilst this is the only published paper investigating the therapeutic application, there are a small number of other studies investigating cleft-type speech characteristics using ultrasound, without testing its therapeutic application. These studies are explored below.

1.5.2.1 Ultrasound and Cleft Palate

As UTI focuses on the unaffected tongue in those with CP, it is unable to directly show the structures of the cleft. However, its value is that it can show real-time lingual movements which are known in some cases to differ from those in children with normal anatomy. Clients are able to view their own tongue in real time and images of the tongue are reasonably easy to explain to clients. Cleland et al. (2013) investigated participants' ability to intuitively *read* ultrasound and EPG images and whether participants were able to interpret images from one tool more so than the other. They found that participants scored above chance in both conditions, suggesting that adults have some degree of intuitive ability to interpret both EPG and ultrasound images. This study used only adult participants, therefore the ability of children to read images is yet to be explored explicitly. As UTI does not require any individualised hardware, it is also relatively inexpensive compared to EPG and it can also be used regardless of a child's anatomy and dentition.

As mentioned in sub-section 1.2.3, retracted articulations are common in the speech of individuals with CP. Gibbon and Wolters (2005) used ultrasound to investigate the speech of an adult with repaired CLP, compared to the speech of a female adult control speaker. They recorded synchronised video-based ultrasound images and audio using a head-probe stabilisation headset. Speech measures included voiceless stops /p, t, k/ and three corner vowels /i, u, a/, in VCV syllables within a carrier sentence, which was recorded four times (12 tokens of each consonant and 24 tokens of each vowel). Using Articulate Assistant (Articulate Instruments, Ltd 2010), they were able to draw splines of the tongue to compare tongue curves qualitatively, annotating and splining vowels at the midpoint and consonants at the point of closure. Results found that in the participant with repaired CLP, vowels were more retracted than those in the control speaker. /i/ and /u/ were closer to /a/. They also found less variation in tongue movements for voiceless plosives in the speaker with repaired CLP, indicated by a smaller standard deviation. While Gibbon and Wolters report that ultrasound has the potential to become a useful tool for investigating abnormal tongue behaviours in speakers with CP, their results were based on a single adult speaker, who they reported had poor image quality due to wires in the jaw and scarring on the tongue. They recorded a limited speech sample, in connected speech with no single word tokens to compare single speech to connected speech tokens, and they did not trial the use of ultrasound as a biofeedback tool for therapy. Gibbon and Lee (2011) suggest that ultrasound is a promising new approach; however controlled studies are required to determine whether ultrasound is a useful tool for speakers with CP.

Bressmann et al. (2011) also investigated compensatory articulations of voiceless velar stops in the speech of individuals with CP. Five participants, with a range of cleft types, including unilateral CLP, bilateral CLP and CPO took part in this study, which included children and adults, aged eight to 23. Within a single recording, using ultrasound with simultaneous audio recordings, participants were asked to repeat five tokens of [aka], [iki] and [uku] through an imitation task. Participants were asked to place their forehead against a headrest and place their chin on the ultrasound probe. Three researchers reviewed the data, through interactive discussion, to document qualitative analysis of the compensatory articulations. Qualitative analysis of the data revealed a variety of typical CP compensatory strategies such as pharyngeal stops and mid-palatal stops. Transcriptions also detected glottal replacement, while UTI also revealed covert articulatory movements which would have been missed using perceptual analysis (Bressmann et al. 2011), such as elevation of the tongue tip and dorsum, and double articulations (for example glottal and velar double articulations). The aim of this study was to establish individual profiles of articulation and to use this knowledge as a benchmark for the client's success in therapy. Only five speakers were recruited; and variability was found between speakers due to the heterogeneity of individuals with CP. In addition, the speech stimuli were limited with only velar targets investigated, because it was reported that these are easily viewable on ultrasound. Alveolar tokens are also easily visible on ultrasound, and are more likely to be associated with placement errors than velars in the speech of individuals with CP, but were not investigated. As Bressmann et al. report tongue tip elevation, it would have been useful to also include a comparison of velar targets with incorrect alveolar tongue tip raising to correct alveolar targets. Furthermore, the speech materials included in this study only include VCV syllables and do not include single words or sentences with the target sound in a range of word positions or vowel environments, with only three vowels chosen. Although articulatory profiles were achieved qualitatively, analysis was based only on three researcher's impressions and there were no quantitative measurements made and no therapy techniques were proposed. While participants rested their forehead against a headrest with their chin on the transducer, this does not fully prevent probe movement. There are no measurements in this paper to account for probe movement. Nevertheless, the explorative study provided additional information regarding the speech of individuals with CP from UTI, such as pharyngeal stops, which would have been unattainable using EPG, suggesting a possible advantage to using ultrasound for individuals with CP.

Although Bressmann et al. did not use quantitative measures, it might be possible to use quantitative analysis of tongue function in speakers with CP, and Zharkova (2013) suggests measures that are designed to normalise across inter-speaker differences in the shape and size of the hard palate and that are usable longitudinally. Two measures (*Dorsum Excursion Index (DEI) and Tongue Constraint Position Index (TCPI)*) quantify the overuse of the tongue dorsum and three measures (*tongue dynamics, variability and separation of tongue curves*) compare sets of tongue curves. It was hypothesised that these measures could be used to compare tongue function pre- and post-therapy and also be used to assess performance against typical speakers to aid the selection of effective treatment (Zharkova 2013). However, Zharkova has not collected any data to test these measurements and it is unknown whether they would be appropriate for lingual data of speakers with CP. Therefore, these proposals remain to be tested, alongside the therapeutic applications of UTI for individuals with CP. Other measures have been used in disordered speakers without CP, for example using paired t-tests to compare two tongue curves (Cleland et al. 2015a; Cleland et al. 2017b). The current study, similar to Gibbon and Wolters (2005) and Bressmann et al. (2011), aims to investigate the compensatory articulations in speakers with CP, using both qualitative and quantitative measurements such as those used in Cleland et al. 2015a and Cleland et al. 2017b.

1.5.3 Visual Articulatory Models

A different set of developments, Visual Articulatory Models (VAMs), generate dynamic videos of articulatory gestures or static images of articulatory target positions within the midsagittal plane (Kroger et al. 2005), and provide a context for lingual movements, for example the hard and soft palates, unlike UTI. While a reference for lingual movements may be useful, VAMs are offline models and do not provide biofeedback. VAMs are geometric models based on static and dynamic Magnetic Resonance Imaging (MRI) data (Kroger et al. 2008). Kroger et al. (2008) propose both a 2D and a 3D model, however they suggest that there is no significant advantage of using 3D over 2D for interpreting the images (Figure 6). They quantified speech recognition rates for mute animations of consonantal and vocalic speech movements generated by 2D and 3D visual articulatory models. Both models were mute and purely visual and were tested using phoneme and feature evaluation. A mimicry test was completed by two groups of children. Each group consisted of eight children (aged five to eight years) with articulation disorders on single sounds but no severe disorders in language development. During the experiment, children were asked to mimic the articulatory movement animations for different speech sounds. The phoneme evaluation showed that 19% of all 3D-model stimuli were produced correctly and 22% of all 2D-model stimuli were matched correctly to the stimuli, indicating that the difference in recognition rate was not significant. Results for the feature evaluation showed that 63% of articulatory features were recognised with the 3D-model and 61% with the 2D-model, therefore there was no significant difference.

Figure 6 2D Visual Articulatory Model (cf. Kroger et al 2008)

Other such models include *Talking Heads* which are virtual clones of a human speaker. The articulatory movements in the Talking Head designed by Badin et al. (2010) are based on the real movements of a single speaker that have been captured using Electromagnetic Articulography (EMA) data (Figure 7). EMA data, however, provides only a small amount of point-based data and the resultant model is extrapolated from these points. These are also based on typical, adult speech and not children's speech and the compensatory articulations discussed above, such as uvular or pharyngeal placement, are not demonstrated on these models, as these are not considered typical productions in English.

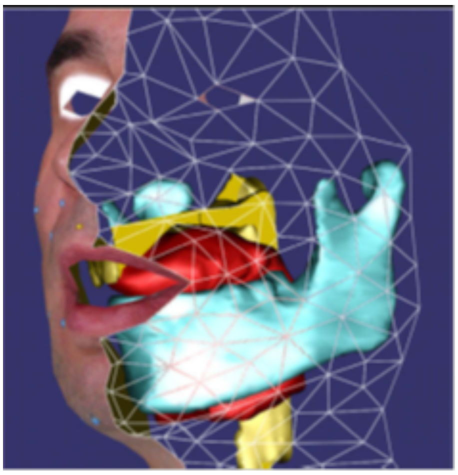
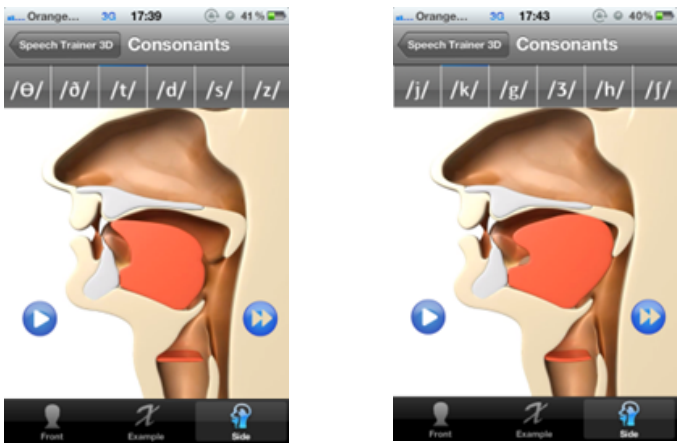


Figure 7 Talking Head (cf. Badin et al. 2010)

These models are not readily available to the public and are expensive. The following sub-section discusses some of the practical reasons why Speech trainer 3D, a VAM available on the iPad, has been chosen over the VAMs presented in this sub-section.

1.5.3.1 Commercially Available Visual Articulatory Models

Commercially accessible articulatory models have recently become available through the iPad (Apple 2012), such as *Speech Trainer 3D* (Smarty Ears, 2011). Speech Trainer 3D's animations are thought to be based on estimations of what might be occurring in the vocal tract rather than instrumental data such as Magnetic Resonance Imaging (MRI) or Electromagnetic Articulography (EMA). Speech Trainer 3D provides animations for both American English consonants from the International Phonetic Alphabet (International Phonetic Association (IPA) 1999) and vowels, in both a front and side view (Figure 8), with only the midsagittal (side) view used in the current study. However, vowels are based on American English and, therefore, are not based on Standard Scottish English. American vowels are different to the Scottish vowels which will be produced by participants of this study, therefore the vowels within Speech trainer 3D will not be used and the focus will be on using Speech trainer 3D for the remediation of consonants.



[t] [k]

Figure 8 Speech Trainer 3D (Smarty Ears 2011)

Although the animations are not always accurate, for example, the lips remain in the same position for both /t/ and /k/ and in the model for /k/, the tongue is shown to touch the uvula; it does create a context for lingual movements, allowing for a detailed description of target positions and movements to clients. It is commercially available; is less expensive than EPG and UTI and finally it does not require specialist equipment or input. However, UTI may have an advantage over VAMs as it shows real-time images of the client's own tongue and provides biofeedback. Speech Trainers 3D's inaccuracy of animations may also be misleading for clients attempting to copy movements.

Over the past few years, the use of technology and apps in Speech and Language Therapy is becoming increasingly popular (Gosnell 2011; Fernandes 2011). It may also be appealing for children to use an iPad, with 40% of children in American in 2013 owning a tablet device (Common Sense Media 2013). In the UK, Ofcom (2014) reported that 65% of children aged three to seven years live in a household with a tablet device, whilst the National Literacy Trust (NLT 2014) found that 72.9% of children aged three to five (from a survey of 1028 children) had access to touch screen devices in the home, including smartphones. Marsh et al. (2015) collected information about preschool (aged 0-5years) children's access to and use of tablet apps in their home, in the UK. They found that 31% of children owned their own tablet. However, there are no studies currently testing whether Speech Trainer 3D is an effective tool in clinic and at present, there is no theoretical basis as to why this tool would be effective. Sub-section 1.5.4 discusses the role of VAMs and UVBF in relation to the theory of motor learning.

1.5.3.2 Visual Articulatory Models and Cleft Palate

To date, there are no publications investigating the clinical applications of Speech Trainer 3D with any population. Other models, based on MRI and EMA data, as mentioned in sub-section 1.5.3., have been designed (Kroger et al. 2005; Kroger et al. 2008; Badin et al. 2010) and tested on participants with typical speech (Badin et al. 2010) and Apraxia of Speech (AoS) and Articulation Disorder in a pilot therapy study (Kroger et al. 2005). Badin et al. (2010) investigated the innate ability of participants to tongue-read, or recover information from visual information on the tongue without priming or learning. If participants were not able to intuitively read tongue images, they then tested their ability to quickly learn how to tongue read. Using a virtual clone of a human speaker, based on EMA data, audiovisual VCV stimuli was presented to participants in audiovisual perception tests in three different conditions - either audio signal alone, audiovisual signal with or without tongue information, or the complete face. With results suggesting the possibility of implicit learning of tongue reading, Badin et al. conclude that models, such as their Talking Head, could be used for the clinical application of children with speech sound disorders, second language learning and perception and production intervention for those with hearing impairment. However, these models are not anatomically accurate for speakers with CLP and the nature of the speech difficulties associated with CP are complex, with retraction to glottal placement and double articulations. The model itself shows the visual information of the tongue movements and lip movements, however there is no information on glottal placement to provide a model to children making cleft type errors, such as glottal replacement or glottal reinforcement.

Kroger et al. (2005) discuss therapy applications for children with AoS (Gotto 2004), Developmental Speech Disorders, i.e. articulation disorders (Albert 2005) and dysarthria (O'Neill 2004). The visual models were used to augment the feedback already provided by the SLT, and were not intended to be used as an independent therapy approach. Gotto (2004) and Albert (2005) did not show any therapy effects of using the VAM as an augmentative therapy tool, due to a variety of influencing factors. However, Kroger et al. (2005) imply that increased rate of recognition suggests an interaction between the participants and the VAM, proposing that speakers may benefit from using the model. Therapy with a larger number of clients is required and would be more beneficial with a wider range of client groups including those with CP.

In addition, Fagel and Madany (2008) use a 3D animated virtual head based on instrumental data as a tool for speech therapy. This was used to teach eight German speakers correct pronunciations of /s/ and /z/. All speakers presented with stigmatismus interdentalis (a pathological production of /s/ and /z/), which is one of the most common SSDs in German children. Although each participant was only given two five-minute learning lessons, there was a significant improvement in six of the eight speakers after one lesson. After the second lesson, only three participants showed further improvement, with only one participant showing a significant difference in production of /s/ and /z/ after the second lesson. This study only included a small amount of data specifically for /s/ and /z/, and learning sessions were very short. This is an error that would not typically be treated in the UK.

Although Speech Trainer 3D is not thought to be based on instrumental data, it is more readily available than the Kroger et al. and Badin et al. models. As there is no theoretical basis for this intervention at present, and evidence for the effectiveness of other models (Gotto 2004; Albert 2005) is limited, it is unknown whether it is an effective therapy tool and therefore, there is a need to test the applications.

1.5.4 The Role of VAMs and UVBF in Motor Learning

Visual feedback tools such as VAMS or UVBF are not therapy approaches in their own right, but are used as adjuncts to therapy approaches (Bacsfalvi et al. 2007), most commonly articulation therapy (Van Riper and Emerick 1984) or motor-based therapy (Preston et al. 2014; Cleland et al. 2015c). Whilst both VAMS, such as Speech trainer 3D, and UVBF can be used for modelling or demonstrating complex articulations, only UVBF provides *biofeedback*. Therefore, it is essential to determine whether it is the VAM element or the biofeedback elements (real-time or delayed feedback) that improves motor learning and in turn whether either tool is effective.

Whilst there is evidence to suggest that people can intuitively read VAMS (Badin et al. 2010), it can be argued that VAMs alone do not appear to be the key ingredient (or the key agent of change) in learning new articulations, due to a number of reasons Cleland and Scobbie (in press). Firstly, while studies have tested the use of VAMs for pronunciation training in second language learning, there is limited evidence for their effectiveness. While Massaro et al. (2008) showed that viewing the lips was useful for teaching [y], for velar and uvular stops [k] and [q] there was no advantage of using the VAM over audio only. Similarly, in a clinical population, Fagel and Madany (2008) showed little effect in treating /s/ and /z/ from interdental production, with a limited number of participants and only one child showing significant differences in their speech after the second lesson. With retraction to velar or uvular placement commonly associated with the speech of children with CP, it could be assumed that using a VAM will have no benefit over traditional motor-based therapy for children with CP, based on the evidence from second language learning (Massaro et al. 2008) and clinical studies (Fagel and Madany 2008) showing no benefit of using VAMs. However, during these studies, participants' learning was not facilitated by an SLT providing KP or KR feedback, which would be typical of motor-based therapy augmented by VAMs.

Children with erroneous articulations, particularly those with motor-based impairments, tend to have an inability to produce the correct articulation on imitation (i.e. the target is not stimulable). Despite surgery, compensatory errors in the speech of children with CP can persist (Chapman and Willadsen 2011), and it is reported that typical phonological processes, such as velar fronting, are more likely to persist longer in children with CP (Chapman and Hardin 1992; Chapman 1993). In this case, it is likely that the auditory system has failed in learning new articulations, irrespective of the surgically repaired structural abnormalities and previous speech therapy using more traditional approaches. This suggests that these children have a lack of understanding of how an articulatory gesture is achieved. Despite this lack of understanding of how to achieve a new sound, there may be some implicit learning in viewing tongue movements, such as in VAMs and UVBF (Cleland and Scobbie, in press).

It is reported that being able to see the lips move during speech (i.e. speech reading) enhances perception (Benoit and Le Goff 1998), with the ability to implicitly learn how to integrate lip information into the perceptual system previously reported, such as in the McGurk effect (McGurk and MacDonald 1976). Similar to speech reading, Badin et al. (2010) suggest that speakers are able to implicitly tongue read a VAM. Evidence of being able to innately lip and tongue read indicates a perception/production link in relation to the theory of mirror neurons (Cleland and Scobbie in press).

Mirror neurons trigger the imitation system of the brain. When a person hears, or sees, an action being performed, they will in turn perform the action themselves. There is evidence to suggest that observing lingual movements using ultrasound activates the premotor and somatosensory cortices (Treille et al. 2014) and that observing a completely novel movement (i.e. an articulatory gesture that is non-stimulable) will generate mirroring activity in the premotor cortex (Cross et al.

2006). When shown the correct production of an articulatory movement, the child's mirror neurons will likely trigger the imitation system to perform the movement themselves, in turn being able to self-regulate and change their own articulations in real-time. Cleland and Scobbie (in press) suggest that the demonstration of correct articulatory movements is a crucial element UVBF. Using a bespoke version of AAA (Articulate Instruments 2012), it is possible to show children videos of correct articulations, in order for the learner to manipulate their tongue in real-time to imitate the tongue movements in the videos. During therapy using UVBF, it is also possible to record and play back the speaker's productions to them. When referring back to the conditions of feedback in motor-learning, delayed feedback is thought to be more useful for learning a new gesture during the practice (Ballard et al. 2010; Murray et al. 2015; McLeod and Baker 2017). This allows the child to reflect on their internal feedback, and therefore, improve their next attempt. This suggests that it is not enough just to watch a new movement, but that there is a need to practice a new motor-plan and be able to self-regulate in real-time, as is the case when using biofeedback.

1.5.5 Summary of Visual Feedback Technologies

There is need for an evidence base for the use of visual feedback technologies, specifically for individuals with CP. A recent Cochrane review highlighted the need for an RCT with the use of EPG for speakers with CP (Lee et al. 2009). Recent studies have analysed the type of tongue movements likely to be found in the speech of individuals with CP, using UTI on a qualitative and quantitative basis (Gibbon and Wolters 2005; Bressmann et al. 2011), however the therapeutic application remains to be tested. Animations, as in Speech Trainer 3D, are yet to be tested within any population.

Whilst there is evidence to suggest that biofeedback has advantage over VAMs in learning a new motor-plan for an articulatory gesture, there are currently no studies that directly compare the therapeutic application of VAMs and UVBF, particularly for speakers with CP. Hence the current study will use Speech Trainer 3D and UVBF, to investigate whether participants are able to improve in their speech outcomes with an offline model such as that in Speech trainer 3D, or whether they require visual biofeedback, when speech errors are residual.

1.6 Summary of Theoretical Background

Most children with CP will present with speech difficulties due to the impaired structure and function of the palate. These speech errors are articulatory in nature but may have a phonological consequence. The active processes found in the speech in individuals with CP are adopted as a means of increasing the sound system. Harding and Grunwell (1996; 1998) and Sell et al. (1994; 1999) have highlighted the need for speech assessment for individuals with CP to include both assessment of articulation and a phonological analysis.

A combination of both articulation and phonological intervention has been discussed in the literature with little evidence to support one technique over the other (Bessell et al. 2013). Although phonetic transcription is deemed gold standard for perceptual assessment of speech in CP (Sell 2005; Peterson-Falzone et al. 2006; Howard 2011), there are issues with the reliability and objectivity of transcriptions (see chapter 3). Visual feedback technologies have been noted as promising tools for assessment (Gibbon and Wolters 2005; RCSLT 2005; Gibbon and Lee 2011; and Bressmann et al. 2011) and in therapy (RCSLT 2005; Gibbon and Lee 2011; Vallino-Napoli 2011; Preston et al. 2014; Cleland et al. 2015c).

EPG has been frequently used for individuals with CP and is recommended by the RCSLT as a clinical tool (RCSLT 2005). The use of ultrasound for investigating cleft-type speech characteristics is relatively recent (Gibbon and Wolters 2005; Bressmann et al. 2011; Zharkova 2013), however its therapeutic application remains to be tested. VAMs provide a context for lingual movements and remain to be tested clinically, with Speech Trainer 3D remaining to be tested in any clinical population. Therefore, there is a need to investigate whether just using schematic, speaker-independent materials to teach lingual movements, as in Speech Trainer 3D, will improve a child's speech, or whether they require the real-time visual biofeedback which is provided using UTI.

This thesis aims to investigate the processes and outcomes of visual tools used in speech therapy for children with a repaired CP. In particular, it investigates the diagnostic use of ultrasound and the effectiveness of UVBF therapy and VAM-therapy, displayed in Speech Trainer 3D (Smarty Ears 2011), through two blocks of therapy which employ a motor-based therapy approach. Three methods of analysis will be presented. Firstly, perceptual scores derived from phonetic transcriptions, including percent target consonant correct, percent consonant correct and intelligibility measures, including inter- and intra-rater reliability measures. Secondly, a perceptual evaluation using multiple phonetically trained listeners. Thirdly, an articulatory analysis of the ultrasound data will be presented, using qualitative and quantitative measurements. Overall, the research undertaken enables a critical evaluation of the potential for a clinical application of UVBF, as well as with testing the applications of one commercially available VAM.

2 Treatment Study

The current study aimed to quantitatively and qualitatively test the effectiveness two therapy tools. Therapy block one comprised of eight one-hour sessions using Speech Trainer 3D (Smarty ears 2011) and therapy block two comprised of eight one-hour sessions using Ultrasound Visual Biofeedback. Both blocks of therapy followed a motor-based therapy approach, similar to Preston et al. (2014) and Cleland et al. (2015c). Three children were recruited, with two participants (Craig and Andrew) receiving therapy. Both standardised and non-standard speech measures were used to evaluate therapy outcomes using narrow phonetic transcription to calculate percent consonant correct, with both inter- and intra-rater reliability calculated. The following section reports the method of the therapy study, with detailed information, methods and results for each individual participant outlined in sections 2.2 and 2.3. A summary of therapy for both participants is provided in section 2.4.

2.1 Treatment Study Method

2.1.1 Study Design

This study used a single-subject multiple-baseline design and compared two treatments using ABACA design (a modification of the ABAB design). The five main phases of the study were: A1) Baseline phase; B) Treatment block one using a Visual Articulatory Model (VAM); A2) Withdrawal period and pre-block two measurements; C) Treatment block two with ultrasound visual biofeedback (UVBF); and A3) Maintenance measurements. Due to the small number of participants (N=2), a cross-over design or group study was not possible.

Single-subject research designs have previously been implemented in speech, language and communication research (McReynolds and Thompson 1986; Brobeck and Lubinsky 2003; Moriarty and Gillon 2006). They provide a quasi-experimental approach to treatment evaluation and effectiveness for either a single subject or a small number of subjects (Backman and Harris 1999). Although some critics say they lack generalisability, this can be established through replication across subjects (McReynolds and Thompson 1986; Backman et al. 1997; Backman and Harris 1999). Brobeck and Lubinsky (2003) note that single-subject designs are particularly useful for discovering whether a given treatment works for a particular participant and out of two or more treatments, which is better for a particular participant. This study aimed to recruit a small number of participants (an initial target of six children) and to test the effectiveness two tools therefore single-subject design was an appropriate choice. Some other advantages of using a single-subject multiple baseline design noted in the literature are that it identifies individual differences between participants; it allows flexibility; it only requires a small number of participants; there is a focus on actual behavioural outcomes; and it has practical clinical application (Irwin et al. 2013).

One essential criteria of the single-subject design is repeated measures during the baseline phases and treatment phases (McReynolds and Thompson 1986; Backman, et al. 1997; Backman and Harris 1999). To comply with this criterion, multiple recordings of untreated words were made during each assessment session and additional recordings of treated words were made during every session of therapy. Treated words were recorded to directly assess the words that had been targeted in therapy and an untreated wordlist was recorded to assess generalisation into words that had not been targeted during therapy. When using a multiple baseline design, reliable and stable baselines of all behaviours must be established (Satake et al. 2008). Stable baseline measurements reduce the likelihood of external events causing change, such as spontaneous maturation. For both participants, repeated measures included the DEAP phonology subtest (Dodd et al. 2002) (see section 2.1.7 below), an untreated wordlist and the intelligibility in context scale (*ICS;* McLeod et al. 2012). These were repeated across six assessment sessions throughout the study (1: Baseline, 2: pre-VAM, 3: post-VAM, 4: pre-UVBF, 5: post-UVBF, 6: maintenance). By repeating these assessments across the baseline and pre-VAM sessions, this allowed us to test whether scores were variable prior to therapy block one. Six comparisons of the scores derived from the repeated assessments were made to measure each phase of the ABACA design and to assess overall improvement. Table 6 gives an overview of the assessment sessions included in each comparison.

Comparison	**Assessment Sessions**					
	1. baseline	2. pre-VAM	3. post-VAM	4. pre-UVBF	5. post UVBF	6. maintenance
Baseline comparison	x	x				
VAM comparison		x	x			
Withdrawal period			x	x		
UVBF Comparison				x	x	
Maintenance period					x	x
Overall Comparison	x					x

Table 6 Summary of Comparisons

As there were repeated measurements and multiple therapy phases within this study, and because the dependent variable (each participant's speech outcome) and independent variables (ultrasound and Speech Trainer 3D app) were already identified, a case study would not have been appropriate. In a case study design, the investigator has little/no control over the results and is considered an observation rather than an experiment. It identifies key themes through qualitative or quantitative data taken from prospective or retrospective studies. A case study provides only a subjective description of the participant's behaviour, rather than an objective measurement provided in single subject design (Backman et al. 1997). A case study can also not test a hypothesis; rather they can only gather information to inform a hypothesis.

2.1.2 Research Questions and Hypotheses

The following research questions are specific to this therapy section only. Research questions and hypotheses for the perceptual evaluation and articulatory analysis will be presented in sub-sections 3.2 and 4.1.1.

RQ1: Will percent target consonant correct (PTCC) of untreated words improve post-treatment after therapy with VAMs and UVBF in the following three comparisons:

- VAM comparison: immediately before (assessment 2) and after (assessment 3) therapy block one using VAMs
- UVBF comparison: immediately before (assessment 4) and after (assessment 5) therapy block two using UVBF
- BL-M comparison: baseline (assessment 1) to maintenance, three months post-therapy (assessment 6)?

H1. PTCC scores will have changed in the following ways in each comparison:

- VAM comparison: PTCC scores will remain relatively stable during therapy block one using VAMs due to previous literature showing no advantage to using VAMS in learning new articulations (for example, Fagel and Madany 2008; Massaro et al. 2008)
- UVBF comparison: PTCC scores will have increased post-therapy block two using UVBF

- BL-M comparison: PTCC scores will have increased by over 20% from baseline to maintenance, showing clinically significant improvement in PTCC (Preston et al. 2014).

RQ2. Will percent consonant correct (PCC) of the DEAP Phonology subtest improve post-treatment after therapy with VAMs and UVBF in the following three comparisons:

- VAM comparison: immediately before (assessment 2) and after (assessment 3) therapy block one using VAMs
- UVBF comparison: immediately before (assessment 4) and after (assessment 5) therapy block two using UVBF
- BL-M comparison: baseline (assessment 1) to maintenance, three months post-therapy (assessment 6)?

H2. PCC scores will have changed in the following ways in each comparison:

- VAM comparison: PCC scores will remain stable during therapy block one using VAM s
- UVBF comparison: PCC scores will have increased post-therapy block two using UVBF
- BL-M comparison: PCC scores will have increased overall from baseline to maintenance

RQ3. Does intelligibility increase post-therapy, as measured by the intelligibility in context scale and parental feedback questionnaires?

H3. The Intelligibility in Context Scale and feedback questionnaires, completed by parents, will indicate improvement in intelligibility post-therapy.

2.1.3 Participants

This study aimed to recruit six children aged 6-16 years, from central Scotland, with a secondary SSD as a result of a repaired CP and included children with all cleft types: syndromic and non-syndromic clefts. Children were excluded from the study if English was not their first language; they had a severe hearing loss; severe language impairment; severe learning difficulty; or if there was any uncorrected visual impairment.

Before recruitment, all ethical issues were considered by the National Health Service (NHS) Research Ethics Committee (REC) and Research and Development (R&D) Office; and the University Ethics Committee.

Participants were recruited through the department of Speech and Language Therapy at a city centre hospital in central Scotland. Specialist Speech and Language Therapists (SLTs) from the CP team were asked to select children who met the inclusion/exclusion criteria. Although the SLTs involved in recruitment had prior knowledge of EPG therapy for CP speakers, they were briefed on the lingual error patterns that were easily viewable in a midsagittal plane prior to recruitment and were therefore suitable for UVBF therapy. Table 7 shows a list of error patterns and whether these errors are imageable on ultrasound or EPG.

Error Pattern	Imageable using EPG	Imageable using Ultrasound
Velar fronting/alveolar backing	Yes (production of velars in back vowels not visible)	Yes
Post-alveolar fronting of /ʃ/ and affricates	Yes (wider tongue groove visible)	Yes (retraction and bunching of the tongue visible)
Palatalisation	Yes	Yes
Retraction to uvular placement	No (EPG palate only goes as far back as the soft palate)	Yes
Retraction to pharyngeal placement	No (EPG palate only goes as far back as the soft palate)	Yes
Sibilant distortions e.g. lateralisation	Yes (tongue groove and lateral bracing visible)	Yes (some information in the coronal view, bunching of the tongue for full closure in midsagittal view)
Double Articulations	Yes (only as far back as velar)	Yes (alveolar/velar/uvular/pharyngeal double articulations visible)
Undifferentiated lingual gestures/overuse of tongue dorsum	Yes	Yes
/l/ errors	Yes (light /l/ visible)	Yes (limited information on lateral bracing)
/ɹ/ errors	Yes (only some information)	Yes (information on variations of /ɹ/: bunched vs. retroflex)
Vowel errors	Yes (limited information for high vowels)	Yes (all vowels imageable)
Stopping of fricatives/affricates	Yes (full closure vs. tongue groove visible)	Yes (limited information)

Table 7 Error Patterns and their imageability with EPG and Ultrasound (Adapted from Wood et al. (2015 p.18) to include errors specific to CP)

Potential participants were first approached by their NHS SLT. For those interested, information packs were distributed and they were asked to contact the researcher directly if they wished to participate. Information packs included a letter to parents/carers for recruitment, information sheets for children and parents/carers and consent forms for children and parents/carers (see appendices in sections 7.1 and 7.2). No identifiable information was obtained until the participants approached the researcher directly. Potential participants were given an opportunity to contact the researcher regarding the study to ask any questions before agreeing to take part. Once participants had made formal contact and had shown interest in participating, they were then invited to attend the University for an initial assessment. Before any data was collected during the initial session, participants and parents/carers were given an opportunity to ask any questions and were then required to give verbal and written consent.

Participants were approached and recruited in a staggered pattern. Seven children were approached initially, of which three responded. One month later, another two children were approached, of which neither responded.

Those recruited were all male aged six to 10 years (6;2, 6;7 and 9;2). Craig (pseudonym), aged 6;2 had a repaired submucous cleft palate, as did Andrew (pseudonym), aged 9;2. Alex (pseudonym), aged 6;7 with repaired CPO withdrew from the study at the baseline phase leaving two participants receiving therapy. For the purpose of this thesis, data from Alex has been excluded. The remaining participants, Andrew and Craig will be discussed in detail in sections 2.2 and 2.3.

2.1.4 Procedure

Each child received six assessment/recording sessions and two blocks of therapy, each with eight one-hour therapy sessions (see Table 8 for the timeline and section 2.1.9 for details on therapy). An initial baseline assessment session was carried out (Assessment 1-week 1), followed by a pre-therapy block one measure (Assessment 2-week 2). The third assessment took place immediately post-therapy block one (week 11) with a five-week break provided before assessment session four took place immediately pre-therapy block two (week 16). Immediately post-therapy block two, a fifth assessment was carried out and finally a sixth, maintenance, assessment was administered three months' post therapy to assess generalisation and retention. Table 8 provides a schedule for assessment and therapy sessions and outlines the speech assessments which were administered within each session.

Baseline Period (weeks 1-2)		Withdrawal Phase (weeks 11-16)		Maintenance Period (week 24 - +3months)	
Overall Comparison (week 1)	**VAM Comparison (intervention weeks 3-10)**		**UVBF Comparison (intervention weeks 17-23)**		**Overall Comparison (+3 weeks)**
Assessment 1 Baseline	**Assessment 2 Pre-VAM**	**Assessment 3 Post-VAM**	**Assessment 4 Pre-UVBF**	**Assessment 5 Post-UVBF**	**Assessment 6 Maintenance**
DEAP Phonology Untreated Wordlist ICS DEAP Articulation Ravens Matrices BPVS	DEAP Phonology Untreated Wordlist ICS CELF4	DEAP Phonology Untreated Wordlist Treated Wordlist ICS Parent and Child Questionnaires	DEAP Phonology Untreated Wordlist ICS	DEAP Phonology Untreated Wordlist Treated Wordlist ICS Parent and Child Questionnaires	DEAP Phonology Untreated Wordlist Treated Wordlist ICS Child Questionnaire

Table 8 Assessment and treatment schedule

2.1.5 Recording Set-up

All assessment sessions and therapy block two were recorded with simultaneous ultrasound, audio and lip-camera. Ultrasound data was acquired using an Ultrasonix SonixRP machine remotely controlled via Ethernet from a PC running Articulate Assistant Advanced software™ (Articulate Instruments Ltd 2012) version 2.14 which internally synchronised the ultrasound and audio data. The echo return data was recorded at 121 frames per second (fps), i.e. around 8ms per frame with a 135 degree field of view (FOV) in a mid-sagittal plane. Recordings and therapy took place in a sound-treated studio with the SLT sitting alongside the participant. Simultaneous acoustic and lip-camera recordings (around 60fps) were also made, using an audio technica 803D clip-on microphone sampling at 22050Hz and a NTSC micro-camera synchronised to the audio. See Wrench and Scobbie (2016) for information on the set up of the Ultrasonix SonixRP.

A bespoke version of AAA, similar to that in Cleland et al. (2015c) was developed to allow the use of the software for therapy. Features included in this version of the software were: saving and calling up target tongue-shapes based on an individual's own productions; superimposing a customised hard palate on the live image; and quick playback of participant's attempts at articulations during therapy or for analysis afterwards. Both speakers were recorded with ultrasound for all speech measures at all six probe time points and the same system was used to provide the real-time visual feedback therapy, enabling a more integrated collection of formal assessment material, but also allowing ad-hoc recording and playback during therapy with the same tools.

A probe stabilisation headset (Figure 9) was used during assessment and therapy sessions to avoid excessive movement of the ultrasound image during the sessions. Only during the second, ultrasound, phase of intervention were the children and treating clinician able to see the ultrasound images during speech and therefore there was no biofeedback available during assessment sessions. To ensure the treating SLT and participants were not able to see the ultrasound images during assessment sessions 1-4, the ultrasound display on the AAA software was minimised.

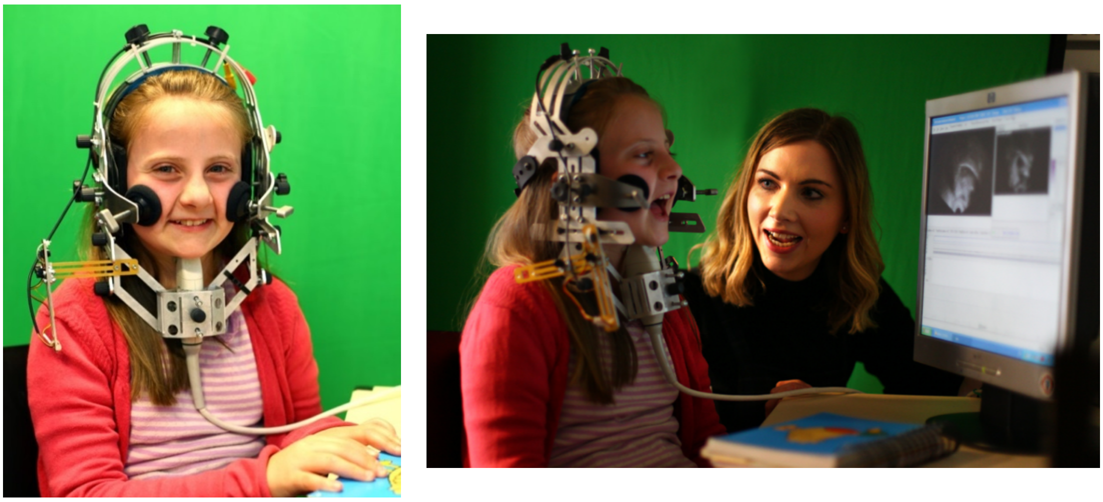


Figure 9 the recording set up illustrating the probe stabilising headset with camera attachment

Figure 9 does not contain images of actual participants in the present study, but from the UltraPhonix Project (2015-2016)[1] with the client's permission. Note that the client-to-screen orientation has been altered to provide a better view of both to the camera in the right-hand image. The 17inch screen demonstrates the bespoke therapy tab, with a static reference in the small box on the right of the screen and the large area on the left of the screen showing the client's real-time image.

A portable ultrasound system was used to provide additional demonstrations and models of tongue shapes for the first two sessions of therapy block two. These were demonstrated by the treating clinician and participant's parents by hand-holding the

[1] UltraPhonix: Ultrasound Visual Biofeedback Treatment for Speech Sound Disorders in Children. **CSO** (Chief Scientist Office) Grant number ETM/402.

transducer under the chin. The probe-stabilising headset was not used for these demonstrations. The portable system used was a B-K Medical Merlin Ultrasound Scanner Type 1101, with an Endocavity End-Fire Transducer type 8561 with a sector of 160 degrees. The transducer frequency employed using this hardware is normally 6.5 or 7.5 MHz. Using this ultrasound system, scanning is performed at a rate of 25-30fps. No recordings were made using the portable ultrasound system.

Therapy block one took place in a university clinic room. Audio data during therapy block one was collected using a Tascam DR100 audio recording system and a boundary microphone. Data was stored on SD cards which were transferred onto university computers for analysis. Therapy block two took place in a sound treated studio for ultrasound recordings and a university clinic room for table-top activities.

2.1.6 Baseline Measures

Language and non-verbal assessments were completed during the two baseline sessions. The *British Picture Vocabulary Scale 3rd edition (BPVSIII;* Dunn et al. 2009) was used to measure participant's receptive vocabulary for Standard English. The Core Language Subtests of the *Clinical Evaluation of Language Fundamentals 4th UK edition (CELF4UK;* Semel et al. 2006) was administered to obtain the Core Language Score in order to determine whether or not participants presented with any language difficulties. The subtests administered for each child were different due to the age differences of Andrew and Craig.

As the study excluded any severe learning difficulty, the *Raven's Coloured Progressive Matrices 1998 edition (*Raven et al. 1998) were used to assess the participant's non-verbal IQ. In this assessment, children are asked to match patterns and select the missing piece of the puzzle using non-verbal problem solving skills.

2.1.7 Speech Measures

2.1.7.1 Formal Assessments

The *Articulation* and *Phonology* subtests of the *Diagnostic Evaluation of Articulation and Phonology (DEAP,* Dodd et al. 2002) were chosen to provide an assessment which sampled all of the consonants of English and also to measure changes to the phonological system during treatment. The Percent Consonant

Correct (PCC) score of the Phonology subtest was used as an outcome measure of the study. Harding and Grunwell (1996) and Morris and Ozanne (2003) report that although the cleft-type errors are articulatory in nature, they may have a phonological consequence on the child's phonological inventory. Therefore, the DEAP phonology subtest was used as a measure of the participants' phonological system, to identify any delayed or disordered phonological processes. As the *Great Ormond Street Speech Assessment Revised* (*GOS.SP.ASS'9;* Sell et al. 1999) had been completed by the CP SLT as part of a routine assessment, this was not administered as part of this study. Results from the GOS.SP.ASS administered before (two-four months) and after (nine months) the participants took part in the study will be included in the individual cases.

2.1.7.2 Target-specific Wordlists

Two target specific wordlists, treated and untreated, were devised in order to identify lingual error patterns using ultrasound, particularly any covert errors unidentified during perceptual analysis. See sections 2.2.2.2.2 and 2.3.2.2.2 for more details on how the targets and words were chosen for the wordlists for individual children.

The untreated wordlists contained thirty-six untreated words and six sentences. These words were never used in the course of therapy, allowing for assessment of generalisation of targets into untreated words. The thirty-six single words were split into groups of twelve words containing the target sound in word initial (WI), word medial (WM) or word final (WF) position. A combination of mono- and polysyllabic words were included. Two words from each group (WI, WM or WF) were then embedded in sentences.

The treated wordlists contained the target speech sound/s in isolation, VC, VCV and CV sequences with a range of vowels, real-words (both mono- and polysyllabic) at single word level, and sentences where appropriate. This was in line with the motor-based therapy hierarchy, similar to that used in Cleland et al. (2017c). The words used in the treated wordlists did not have accompanying pictures.

An additional wordlist was devised for Andrew (see subsection 2.2.2.2.2), which contained the target speech sound with distractors and minimal pairs where appropriate. Each target was presented in isolation, VC, VCV and CV syllables before being presented in WI, WM and WF positions in real single words. The target sound was also presented in clusters in WI, WM and WF positions. An additional list was also created for Craig, containing minimal pairs and distractors for his velar target, however due to time constraints and Craig's level of motivation, this was not recorded for Craig. The analysis of this wordlist is separated from the untreated words, as some of the words in this list were used during therapy and it was not included in inter- or intra-rater reliability measures.

2.1.7.3 Intelligibility Measure

The Intelligibility in Context Scale (*ICS*; McLeod et al. 2012) was completed by one parent (Mother) during all assessment sessions apart from assessment two (pre-VAM). The ICS is a seven-item questionnaire used to measure the degree to which participants are understood by a range of listeners (parents, immediate family, extended family, friends, acquaintances, teachers, and strangers). It is a measure of functional intelligibility which is easily administered and is a valid and reliable measure of children's intelligibility (McLeod et al. 2012). Scores from the ICS were rated on a five-point scale (1=never understood; 2=rarely understood; 3=sometimes understood; 4=usually understood; and 5=always understood), with regards to each child's intelligibility when speaking to listeners of varying degrees of familiarity, such as immediate and other family members, teachers and strangers. Questions read, for example, "Do you understand your child?" through to "Do strangers understand your child?".

2.1.8 Post Therapy Questionnaires

Three questionnaires were devised and were completed by participants and parents. The therapy outcome questionnaire for parents (see appendix in section 7.4) comprised of 11 questions alongside the completion of the ICS. Questions two-to-four asked parents to circle their child's target sounds and what context they were practiced in (single words, syllables, whole words, sentences, key vocabulary). As parents sat in the therapy sessions, they were able to observe therapy targets and what level of the therapy hierarchy their child was working at. Questions five-to-ten focused on their child's progress during therapy. Parents were asked to rate the improvement in their child's speech and phonological awareness in the questions. Questions five and six are outlined below as an example. For all questions, refer to appendix 4.

"Please rate your child's progress with their speech since enrolling in the project:

greatly improved, moderately improved, not improved, slightly deteriorated, greatly deteriorated

After treatment, my child's awareness of speech sounds has:

greatly improved, moderately improved, not improved, slightly deteriorated, greatly deteriorated".

Question 11 asked parents: "Please comment on whether or not you think using Speech Trainer 3D/Ultrasound (delete as appropriate) has made it easier for your child to achieve his speech therapy goals". Question 12 asked parents for any other comments.

The therapy outcome questionnaire for children (see appendix in section 7.3) asked questions regarding their views on the treatment tools, e.g. "What did you think about using the ultrasound/iPad?" and "What was the best/worst bit about using this tool?". They were also asked questions related to intelligibility, using the same five-point scale as the ICS, for example "How often do you think your parents understand you when you speak?" and "When you talk to new people, how often do they understand you when you speak?". The three-month post therapy questionnaire for children (see appendix in sub-section 7.5) asked for the participant's preference of tool and why.

2.1.9 Therapy

Therapy was provided by a qualified speech and language therapist (SLT, the author). The first block of therapy used the iPad app "Speech Trainer 3D" (Smarty Ears 2011) as a visual articulatory model (VAM) and the second block of therapy used UVBF. Similar to Preston et al. (2014), sessions were evenly distributed as much as possible to ensure that both ultrasound and non-ultrasound therapy tasks were applied. Preston et al. (2014) use a timer for 13 minute periods with and without ultrasound. Typically, in the present study, the first 30 minutes of each session focused on using either Speech Trainer 3D or Ultrasound, as appropriate, alongside a motor-based therapy approach (see below) and the second 30 minutes on traditional table-top activities focusing on the same target sound to build in generalisation early on within therapy sessions. An eclectic therapy approach, which is described by Joffe and Pring (2008) as a favourable approach by clinicians, was used during the latter 30 minutes and incorporated phonological approaches such as minimal pairs (Barlow and Gierut 2002; Baker 2010) and non-word auditory discrimination and production based on the psycholinguistic model (Stackhouse and Wells 1997). The therapy targets were based on discussion with the Specialist CLP SLT prior to participants beginning therapy. During the course of therapy, Craig also received Speech and Language Therapy by the Specialist CLP SLT to target additional sounds, biliabial stops /p/ and /b/ and fricatives /f/ and /s/, which were not treated within this study. Craig continued to receive therapy throughout the course of the study as it was felt unethical to withdraw any other treatment due to multiple speech errors. Andrew did not receive any other Speech and Language Therapy whilst taking part. Individualised therapy is discussed in sections 2.2.2.4 and 2.3.2.3. Below is an overview of the general therapy procedure.

2.1.9.1 Therapy Approach

Due to the articulatory nature of speech disorders in CP, an articulation or motor-based therapy approach is the preferred approach for individuals with CP (Peterson-Falzone et al. 2010). The therapy technique used in this study followed the principles of motor learning through an articulation therapy approach, such as those in Preston et al. (2014), Hitchcock and McAllister Byun (2015), and most closely, the protocols of Cleland et al. (2015c) developed during the ULTRAX project (2011-2014)[2].
Acquisition was measured during therapy sessions through SLT perceptual judgement in therapy block one. During therapy block two the ultrasound also allowed the treating SLT to assess whether the participants achieved a correct or incorrect tongue shape for their target sound and whether they presented with any covert errors, for example double articulations, and therefore allowed the SLT to measure acquisition of a new gesture. As practice improves accuracy, performance

[2]Ultrax: Real-time tongue tracking for speech therapy using ultrasound. **EPSRC** (Engineering and Physical Sciences Research Council) Healthcare Partnership research grant EP/I027696/1.

should increase within sessions over repeated trials. Practice and repetition was implemented and performance was monitored during individualised therapy activities each week by making audio and video recordings. The treating clinician also did live scoring during each session. However, learning does not occur through performance alone and must be observed through generalisation and retention (Maas et al. 2008; Schmidt and Lee 2005). Both generalisation and retention reflect motor learning (Sjolie 2015). Therefore, generalisation was assessed through the untreated wordlist in both single words and sentences and retention was measured through the post-VAM and post-UVBF assessments and the three-month maintenance session. This allowed the comparison of the effects of acquisition and learning in therapy using an off-line visual articulatory model (VAM) and ultrasound visual biofeedback (UVBF). Feedback included Knowledge of Performance (KP) and Knowledge of results (KR) feedback. Feedback began with high intensity which reduced to low intensity as therapy progressed. Feedback was given either concurrent with each attend at a target through biofeedback, immediately after through auditory or visual biofeedback, or delayed feedback. The bespoke version of AAA used for therapy allows you to record a child's attempt and play it back to them. For the delayed feedback, numerous attempts (around 10) were recorded without providing any KP or KR feedback and then videos were watched immediately after by the child and SLT. This allowed for discussion of the errors together and for the child to use their internal feedback system to self-regulate and improve on their next attempt. Table 9 summarises the practice and feedback conditions used in the current treatment study, in line with the principles of motor learning.

		Conditions used in the current study
Practice condition	Practice Amount	Large practice amounts (>100 trials per session)
	Practice Distribution	Distributed (eight one-hour sessions once weekly)
	Practice Variability	Variable and Constant (linguistic variability)
	Practice Schedule	Blocked (each target ten times) and randomised (during table top activities)
	Practice Complexity	Increasing complexity with a hierarchy similar to Cleland et al. (2017c)
	Practice Fraction	Simplification, segmentation and shaping
	Practice Accuracy	Errorful learning
	Attentional Focus	Internal and external with VAM and UVBF. External during table top activities for retention and generalisation
Feedback Condition	Feedback Type	Knowledge of performance (KP) and knowledge of results (KR)
	Feedback Frequency	High frequency during earlier sessions and low frequency during later sessions
	Feedback Timing	Concurrent, immediately after and delayed (including playing own recording back to participants to reflect on internal feedback and self-monitor)

Table 9 Practice and Feedback Conditions Used in the Treatment Study

2.1.9.2 Therapy Block One – Visual Articulatory Model

VAMs, used in therapy block one, were presented in a mid-sagittal plane on an iPad3 (Apple 2012) via the app Speech Trainer 3D, developed by Smarty Ears (2011). The iPad was chosen as a tool due to its current appeal in society. Mobile apps, particularly those on iDevices (Apple 2012) are becoming increasingly popular and are now becoming a key tool for Speech and Language Therapists (Gosnell

2011). In 2011, at least one in five SLT clients used a handheld device (Dunham 2011). This particular app was selected for this study due to its commercial availability, low cost at £7.99 and simplicity of use. Speech Trainer 3D shows a midsagittal view of the vocal tract, including a reference for tongue movements in relation to other areas of the vocal tract, unlike ultrasound which shows only the tongue surface. However, to our knowledge, it is not based on articulatory data and is not always accurate, for example the video for /k/ is out of sync and the velar images show closure in the uvular region. Figure 10 shows an example of inaccuracies of velar images. The creators of Speech Trainer 3D were contacted prior to the start of this study to confirm whether the animations were based on articulatory data, with no response.

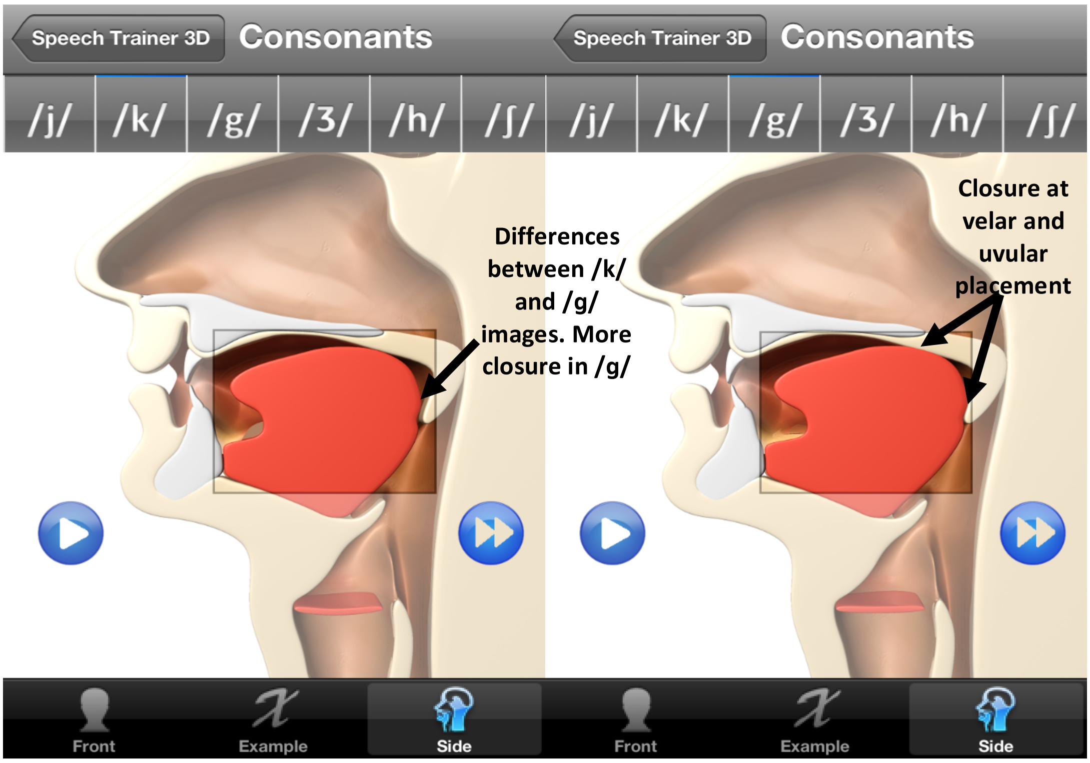


Figure 10: Inaccuracies in velar productions in Speech Trainer 3D (Smarty Ears 2011) (N.B. images are produced under fair use, without permission). Figure on the left is /k/, Figure on the right is /g/.

The first and second therapy sessions for each child focused on using Speech Trainer 3D to demonstrate the parts of the vocal tract and to demonstrate and label the gestural components of the sounds of English. During these sessions, auditory discrimination tasks and visual discrimination tasks using still frames from the Speech Trainer 3D model were implemented. Production practice was individualised, but followed a motor-based approach similar to Preston et al. (2014) or a traditional

Articulation Hierarchy (Van Riper and Emerick 1984). As both children were not stimulable for the target articulation as identified through assessments, therapy began by using models of phonetically similar segments (see Cleland et al. 2015c) that the children were able to achieve (see sections 2.2 and 2.3 for details). Audio recordings were made during each session and were played back to participants to monitor their own progress and to identify their own errors. When listening back to the delayed auditory feedback of the audio recordings, participants were asked to use their own perceptual judgement, alongside the SLT's perceptual judgement, to decide whether their production was correct or incorrect. Using Speech Trainer 3D as an aid to view the animations in correspondence to their correct and incorrect productions of their errors, the participants would identify the sound that they heard when listening back to their productions. If they identified that they produced the sound incorrectly, they were asked to identify which sound they thought they were producing using the animations on Speech Trainer 3D. They were then asked to watch the animation of the target sound and describe the changes in lingual gestures they had to make in order to achieve the correct production of a sound.

2.1.9.3 Therapy Block two – Ultrasound Visual Biofeedback (UVBF)

As in Cleland et al. (2015c), the first therapy session for each child focused on learning to associate the movement of the ultrasound image on the screen with the movement of their own tongue by demonstrating tongue shapes already in their inventory. During this first session, ultrasound was also used as a visual articulatory model, with the SLT demonstrating tongue shapes on a portable ultrasound system, and ultrasound images were compared to those in Speech Trainer 3D. Children were asked to match dynamic ultrasound videos to the dynamic videos in Speech Trainer 3D by scrolling through the Speech trainer 3D app and identifying the target tongue shape that matched the ultrasound images. During the first two sessions, sounds were also modelled to them by the treating SLT using a portable ultrasound system. Like the first block of therapy using Speech Trainer 3D, production practice was individualised and followed a motor-based approach.

Similar to Cleland et al. (2015c), and to the first block of therapy, therapy began with shaping new articulations from phonetically similar segments. Since both children were able to produce a perceptually acceptable target within the first block of therapy, in at least one condition (e.g. VC or CV), their own best attempt from the pre-therapy assessment recording was used as a target tongue-shape. For both children, the same target was chosen for therapy block two. See sections 2.2 and 2.3 for individualised therapy plans and target rationales. Children were encouraged to watch their own recordings and to listen to audio recording to provide delayed feedback.

2.1.10 Data Analysis

All speech assessments were phonetically transcribed using IPA symbols (IPA 1999) and the extIPA Symbols for disordered speech (IPA 1999). As the data were audio recorded, transcriptions were not live. There was no restriction on the number of times each token was listened to. A narrow phonetic transcription of speech from the first block of therapy was performed pre-therapy block one using the acoustic and lip-camera data. Ultrasound data was not viewed prior to therapy block one so as to not influence therapy using Speech Trainer 3D with no biofeedback. Post therapy block one, acoustic, ultrasound and lip-camera data were used for narrow phonetic transcription.

From the transcriptions, a correct/incorrect score of 1 or 0 was used to obtain a percent target consonant correct (PTCC) score from the untreated wordlists. A percent consonant correct (PCC) score was obtained from the DEAP phonology subtest following manual instructions. Using the Intelligibility in Context Scale, a total intelligibility score and an average total score was obtained. These three scores were compared across the six assessment time-points to make six comparisons (Table 10).

Comparisons	Assessment Sessions					
	1. baseline	2. pre-VAM	3. post-VAM	4. pre-UVBF	5. post UVBF	6. maintenance
Baseline comparison	x	x				
VAM comparison		x	x			
Withdrawal period			x	x		
UVBF Comparison				x	x	
Maintenance period					x	x
Overall Comparison	x					x

Table 10 Overview of Comparisons

Articulatory analysis of the ultrasound data will be explored further in chapter 4 and results will be presented individually for each participant (see section 4.2).

2.1.10.1 Intra-rater reliability

A second narrow transcription of the data derived from the untreated wordlist was completed three years after it was first transcribed to provide intra-rater reliability. As with the inter-rater reliability, all 36 single-words were re-transcribed. As only single words were included in the multi-listener perceptual study (see chapter 3) and for articulatory analysis (see chapter 4), only single words were used for both inter- and intra-rater reliability and connected speech samples were not included. PTCC was again calculated at each time point. Intra-rater agreement of the PTCC scores was then calculated for each assessment using a Cohen's Kappa (Cohen 1960). The Cohen's kappa is a measure of agreement which adjusts the observed proportional agreement to take account of the amount of agreement which would be expected by chance. The interpretation of kappa, after Landis and Koch (1977) is as follows:

- <0.20 Poor
- 0.21-0.40 Fair
- 0.41-0.60 Moderate
- 0.61-0.80 Good
- 0.81-1.00 Very good

A percentage of transcription reliability is presented across all six therapy sessions. This was scored using a three-point system to rate the equivalence of transcriptions on a token-by-token basis:

0. Different: for example, if at one point an alveolar nasal target was transcribed as correct and at the other time-point was transcribed as a velar nasal; or if at one time point a velar was transcribed as an alveolar stop but transcribed as an alveolar-glottal double articulation at the other time-point, these were considered different.
1. Almost equivalent: This includes *functional equivalence*, i.e. essentially equivalent phonetic transcriptions of a target behaviour that uses alternative symbolisation; and *near functional equivalence*, i.e. nearly equivalent phonetic transcriptions of a target behaviour in terms of place and manner features. For example, if /k/ is transcribed as [k] in the first set of transcription and [k^h] in the second set, or if /ŋ/ is transcribed as [ŋ] in the first set and [ŋ̃] in the second set.
2. Identical.

2.1.10.2 Inter-rater reliability

Broad phonetic transcriptions were carried out by two further experienced phoneticians using audio data only, in order to compare percent target consonant correct (PTCC) scores derived from the untreated wordlists. All 36 single words from the untreated wordlists from all six assessment sessions were transcribed. Audio data of the 36 single words was exported from AAA into PRAAT version 5.3.57 (Boersma and Weenink 2013). Individual words were edited from longer recordings (three words per recording) hence silence was included either side of each word where possible. Each single word was saved as an individual WAV. file.

Listeners who were transcribing data for inter-rater reliability measures were informed of the therapy targets for each participant and were instructed to transcribe only the target sound within each of the 36 words using broad transcription (e.g. /n/ for Andrew and velars for Craig). There was no restriction to the number of times the listeners could listen to each target. The order of the sessions was randomised so that the listeners were blinded to information about which time point the data was recorded in.

From these transcriptions, PTCC was calculated at each time point by giving each token a score of 1 (correct) or 0 (incorrect). PTCC scores were compared to the PTCC scores derived from the author's first set of transcriptions as the second set of transcriptions for intra-rater reliability were carried out after inter-reliability measures were completed. Where the author's transcriptions were narrow and the other transcriptions were broad, they were compared in terms of place of articulation, voicing and manner and the use of diacritics was discounted. Due to these differences in transcription (i.e. broad vs. narrow), equivalence scores, as used for intra-rater reliability, were not possible across three transcribers. The two additional transcribers did not use diacritics, therefore there were no "almost equivalent" options and transcriptions were either the same or different.

Listener agreement of PTCC scores was calculated using a Fleiss' Kappa (Fleiss 1981). A Fleiss' Kappa can be used to measure agreement among listeners and transcribers (see below). Fleiss' Kappa results can be interpreted in the following way: <.40 = Poor agreement; .60–.74 = Intermediate to good agreement; ≥.75 = Excellent agreement (Fleiss 1981).

2.1.11 Summary of Treatment Method

The treatment study provided Andrew and Craig with two blocks of therapy, with VAMs and UVBF, each block containing eight one-hour therapy sessions. Participants also attended six assessment sessions. The DEAP phonology subtest and an untreated target specific wordlist were used as repeated speech measures. The following two sections will provide information on specific materials, PTCC and PCC results and a clinical discussion for each participant. Articulatory Results presents the results of an articulatory analysis of the ultrasound data.

2.2 Andrew

2.2.1 Background Information

2.2.1.1 General clinical profile

Andrew (pseudonym), aged 9;2, has a repaired submucous cleft palate. Andrew was referred to the specialist CP CLT service at age 2;11 to assess palatal movement. On referral, Andrew only presented with consonants /m/ and /n/. No other consonants were stimulable. He had undergone grommet insertion twice due to Otitis Media with Effusion and fluctuating conductive hearing loss and had his tongue tie clipped as a baby. Andrew had received therapy from the community SLT and from the CLP specialist SLT. See below for details on his multi-disciplinary input. At the time of referral to the project, the CLP specialist SLT reported that Andrew was backing the alveolar nasal /n/ to a velar nasal stop. He had normal resonance with no audible nasal emission, however he did have inconsistent nasal turbulence. He had adequate oral pressure for high pressure consonants. His SLT requested alveolar nasal /n/ as a therapy target on the project. During the course of the project, he did not receive any additional treatment from his community or cleft specialist SLT.

2.2.1.2 Underlying Condition

Andrew has a diagnosis of hemifacial microsomia with unilateral microtia on the right side, resulting in a mild unilateral conductive hearing loss. Hemifacial microsomia is a congenital disorder that affects the lower half of the face, most commonly the mandible, the ears and the mouth. Fan et al. (2005) suggest that there is a concurrence between hemifacial macrosomia and CLP. In their study, 20 out of 198 participants with hemifacial macrosomia also had CLP (10%), suggesting an aetiopathalogic link. Andrew's speech errors, such as retraction, are characteristic of his submucous CP rather than his associated hearing impairment caused by unilateral microtia.

2.2.1.3 Chronology of Andrew's MDT interventions and diagnosis

At the point of referral to the CLP SLT service, it was reported that Andrew also had grommets inserted. He presented with very few consonants, /m/ and /n/ only, and a range of vowels. He was therefore referred to the CLP specialist service for assessment of palatal function. Table 11 summarises the input Andrew received between referral to the CLP service and the referral to the current project at age 9;2. The key information from Andrew's chronology is summarised below. Information is taken from case-notes from the CLP specialist SLT and is organised into SLT input, MDT instrumental assessment and surgery.

Age (years;months)	SLT Input	MDT Input	Surgery
2;11	Referred to CLP SLT service		
3;0	CLP SLT Ax		
3;0	CLP SLT Tx and Ax		
3;10		MDT Cleft Clinic Ax	
3;11	Community SLT Tx		
4;2	Community SLT Tx		
4;8	Community SLT Tx		
4;9		MDT Consultant Report	
5;0	CLP SLT Ax		
5;8	Community SLT Tx		
5;10		MDT Consultant Report	
5;12		MDT Cleft Clinic Ax	
6;0			Submucous cleft palate with associated velar palsy repaired in January 2010. Radical dissection and retro positioning on the functioning left levator veli palatine and explored the right (immobile) side to identify any levator tissue in order to reconstruct the levator sling with radical mobilisation of the left (functioning) levator.
6;1		Hearing Ax	
6;4		MDT Cleft Clinic Ax	
6;7	Community SLT Tx		
6;9		MDT Consultant Report	
7;1		Hearing Ax	
8;2		MDT Cleft Clinic Ax	
8;5			Right unilateral Hynes Pharyngoplasty to treat asymmetric VPD related to hemifacial microsomia and right velar palsy
9;0		MDT Consultant Report	
9;3	Referred to University research project		

Table 11 Summary of Andrew's input from referral to CLP SLT service to referral to the current project

2.2.1.3.1 Initial Assessment with the CLP Specialist Team

When Andrew was first referred for assessment by the specialist team at age 2;11, assessment indicated that he had good comprehension which had improved after grommet insertion. His phonological system was extremely limited, using mostly vowels with a range of intonation patterns. He had difficulty planning and processing speech sounds and had possible VPD with possible low tone or lack of efficient velar function. At age 3;0, a CLP SLT assessment indicated that he had difficulty with oromotor tasks such as lip rounding and imitation of lingual movements. At this stage, it was reported that he was too young to assess palatal function using lateral videofluoroscopy or nasendoscopy. He was referred to community SLT and to the consultant CLP and plastic surgeon aged 3;0.

2.2.1.3.2 SLT Input: Assessment and Therapy

Andrew received input from both the CLP specialist SLT and a community SLT. Initially, at age 3;0, the specialist SLT targeted oromotor movements in 1:1 sessions. Between the ages of 3;11 and 4;08, therapy from the community SLT targeted phonological awareness and production of fricatives. Input was delivered in a small group. By age 5;0, he was able to produce a range of high pressure consonants, [p b t d k g s f]. He was able to produce [t͡ʃ] and [d͡ʒ] in single words but not spontaneously. It was reported by the CLP SLT that placement for /s/ was variable and was retracted to palatal placement inconsistently. He presented with inconsistent nasal turbulence with some hypernasality. Andrew continued to receive therapy from his community SLT, to achieve oral pressure for high pressure consonants and to target /l/. Therapy was less frequent between the age of 5;8 and the time of referral to the current project. He was regularly assessed by the CLP specialist team.

The GOS.SP.ASS'98 (Sell et al. 1999) was completed by the CLP SLT four months prior to Andrew starting on the current research project. Table 12 shows a summary of the GOS.SP.ASS'98 completed by the CLP SLT prior to Andrew starting therapy on the current project. Green shading indicates consonants present in Andrew's inventory in both SIWI and SFWF positions. Prior to beginning the study, Andrew retracted his alveolar nasal to velar placement. VP friction was evident on SFWF /t/, SIWI /s/ and SIWI /d͡ʒ/, along with lateralisation on SFWF /ʃ/ and /d͡ʒ/ and palatalization on SIWI /ʃ/.

	Labial					Alveolar						Post-Alveolar			Velar			Glottal	
	m	p	b	f	v	n	l	t	d	s	z	ʃ	ʧ	ʤ	ŋ	k	g	h	θð
SIWI						ŋ		t		s͌		ʃʲ		ʤ͌			g		
SFWF						ŋ		t͌		s		ʃˡ		ʤˡ			g↑		

Table 12 Andrew's GOS.SP.ASS'98 Consonant Production pre-study (green shading indicates consonants present in Andrew's inventory)

Whilst there is evidence of VPD on more than one high pressure consonant, resonance was not the focus of the current study. Therefore, /n/ was requested by the CLP SLT as the therapy target due to errors with lingual placement. Due to assessment results (see below) and the background information provided on Andrew, it was agreed that /n/ would be the therapy target.

2.2.1.3.3 MDT Input: Assessment and Surgery

Andrew had assessments, mostly annual, with the MDT cleft team. He attended assessments from the SLTs, audiologists and the surgical team. He had intra-oral assessments and instrumental assessments with X-Ray, 2D videofluoroscopy, lateral videofluoroscopy and nasendoscopy to assess his palatal function.

Andrew received surgery at age 6;0 to repair his submucous cleft palate and associated velar palsy. Surgery involved a radical dissection and retro positioning on the functioning left levator veli palatine and explored the right (immobile) side to identify any levator tissue in order to reconstruct the levator sling with radical mobilisation of the left (functioning) levator.

Andrew received secondary surgery aged 8;5. The surgical team performed a right unilateral Hynes Pharyngoplasty to treat asymmetric VPD related to hemifacial microsomia and right velar palsy. Andrew's speech was reviewed two days post-surgery. He presented with mild and inconsistent hyponasal resonance with no hypernasal resonance, mild and inconsistent nasal turbulence with no nasal emission and no passive cleft speech characteristics. It was noted, however, that he had lateralization of /ʃ/ /t͡ʃ/ and /d͡ʒ/ inconsistently and that he was retracting /n/ to velar placement. Two months post-surgery, Andrew had normal resonance, no audible nasal emission and inconsistent nasal turbulence. He had no weak or nasalised sounds, however retraction of alveolar nasals to velar placement persisted post-surgery.

2.2.2 Method

In the current study, Andrew received six assessment sessions and two blocks of therapy, each with eight one-hour therapy sessions. See section 2.1 for the general procedure and recording set-up. Details of therapy will be described below.

2.2.2.1 Language and Non-Verbal Measures

Andrew attended two one-hour long baseline sessions prior to therapy block one. Language measures (see section 2.1.6) were completed during these two sessions. Andrew's receptive vocabulary measured in the borderline-normal range with a standard score of 78 on the BPVS-III (Dunn et al. 2009). In contrast, his language score was in the normal range, with a standard score of 99 in the CELF-4UK core language subtests (Semel et al. 2006) Table 13 shows a breakdown of individual subtests from the CELF-4UK. Non-verbal IQ was also tested using the Raven's Coloured Progressive Matrices (Raven et al. 1998), showing that Andrew's score was in the 75th percentile (Grade II "definitely above average intellectual capacity").

Subtest	Raw Score	Scaled Score	Scaled Score Points +/-	Confidence Interval 90% level	Percentile Rank	Percentile Rank confidence interval
Concepts and Following directions	50	11	2	9 **to** 13	63	37 **to** 84
Recalling Sentences	61	9	1	8 **to** 10	37	25 **to** 50
Formulated Sentences	46	10	2	8 **to** 12	50	25 **to** 75
Word Classes-Total		10	1	9 **to** 11	50	37 **to** 63

Table 13 CELF4 Scores for Individual Subtests

2.2.2.2 Speech Measures

A range of formal and targeted wordlists were used to measure speech outcomes. Table 14 gives a summary of the speech measures used in each assessment session. Details of each speech measure are outlined below. All formal assessments and target-specific wordlists were recorded with simultaneous ultrasound.

	Pre-Study	Baseline	Pre-VAM	Post-VAM	Pre-UVBF	Post-UVBF	Maintenance	Post-Study
Formal Speech Measures								
GOS.SP.ASS'98	X							X
DEAP Phonology		X	X	X	X	X	X	
Target-specific wordlists								
Untreated /n/		X	X	X	X	X	X	
Treated /n/				X	X	X	X	
Additional Alveolar Wordlist		X	X	X	X	X	X	
Questionnaires								
ICS		X	X	X	X	X	X	
Parent-Questionnaire				X		X		
Child-Questionnaire				X		X	X	

Table 14 Summary of Speech Measures

2.2.2.2.1 Formal Speech Measures

The DEAP (Dodd et al. 2002) was chosen to provide a PCC score and to measure changes in Andrew's phonological system over the course of treatment. The DEAP Phonology subtest was recorded at all six assessment time-points (Assessment 1-6) using synchronised ultrasound (see section 2.1.5), audio and lip camera data. As the CLP SLT requested /n/ as the therapy target, the number of alveolar nasals in the DEAP was checked by counting the number of alveolar nasal targets (n=8). Due to the small number of tokens, further wordlists were created to investigate Andrew's production of /n/, which contained both treated and untreated syllables and words.

2.2.2.2.2 Target-specific Wordlists – materials and protocol

2.2.2.2.2.1 Untreated Wordlist

As described in Treatment Study Method, the assessment protocol also included an "untreated" wordlist to assess the specific therapy target in words that were not treated during therapy in order to assess generalisation into untreated words. PTCC scores were calculated at each assessment time point. The untreated /n/ wordlist consists of 36 single words containing 39 tokens of /n/ in (mostly singleton) word initial, (mostly intervocalic) medial and (mostly singleton) final positions in a variety of vowel environments (see Table 15). A range of picturable monosyllabic and polysyllabic words were used. Each word was elicited through a picture naming task. Pictures were presented to Andrew in blocks of three. If Andrew was unable to name the word spontaneously a semantic cue was provided followed by direct imitation if this failed to elicit the desired response. Photographs from Google images were used for picture naming tasks. Six sentences were also recorded which included a sample of the 36 single words in connected speech. Sentences were elicited through an imitation task. Table 15 presents the wordlist with sentences, organised into word positions, vowel environments, clusters and sentences. The vowel choices are appropriate for the accent of the child.

	Vowel	Untreated Word
WI (12)	/i/	neeps*, kneeling
	/ɛ/	necklace
	/a/	nachos, nappy
	/ɔ/	knot
	/o/	gnome, notebook, nose
	/ʌ/	nuts
	/ɪ/	knitting, nibbling
WF (11)	/i/	green
	/a/	can, snowman
	/ɔ/	leprechaun
	/o/	phone, bone
	/ɪ/	garden, violin, medicine, curtain
	/ə/	skeleton
WM (12)	/ʉ/	tuna
	/ʌ/	sunny, funny
	/a/	banana, animals, piano
	/ɪ/	dinner
	/aʊ/	brownie
	/ae/	dinosaur
	/ə/	lemonade, banana, vanilla
Clusters (4)		snowman popcorn onions
Sentences (25)		Nadine's Nana was knitting her a nice jumper
		Nina's notebook is very neat
		Neil and Noah made a snowman
		Nora answered the phone
		Nick is having chicken for dinner
		Benjamin had a banana for lunch

Table 15 Untreated /n/ wordlist organised into word positions, vowel environments, clusters and sentences. Brackets indicate the number of tokens of /n/ in each environment.

The tokens of /n/ in these sentences were then scored separately from the single words to obtain a separate PTCC score. Prior to starting therapy, Andrew's PTCC score for single words on the untreated wordlist was 5% at baseline (Assessment 1) and remained relatively stable at 8% in the pre-VAM session (Assessment 2).

2.2.2.2.2.2 Treated wordlist

A treated /n/ wordlist provided materials for use during intervention (see Table 16). This wordlist consists of words containing /n/ in WI and WF position. Where appropriate, /n/ words were paired with an /ŋ/ minimal pair and recorded in order to show Andrew the similarities and differences in tongue shapes in WF positions during therapy. /n/ and /ŋ/ cluster minimal pairs were also recorded. These included /nz/ /ŋz/ and /nd/ /ŋd/ clusters. This wordlist was recorded post-VAM, pre-UVBF, post-UVBF and maintenance. Treatment materials were modified throughout therapy block one in response to developing client needs. The wordlist was not recorded at baseline or pre-VAM because it was tailored to suit treatment needs and modified during therapy block one. Therefore, there was no pre-post comparison made for therapy block one.

	Vowel	Treated Word	Minimal Pair
WI (13)	/i/ /e/ /ɛ/ /ʉ/ /ɪ/ /aʊ/ /ae/ /ɔɪ/ /juː/	neat name, nail net, neck noose nib now nine, night, knife noise new	
WF (7)	/a/ /ʌ/ /ɪ/ /ae/	fan, ran sun pin, thin, win nine	fang, rang sung ping, thing, wing
Clusters (6)		fans, band, wins buns, pinned, tonnes	fangs, banged, wings bungs, pinged, tongues
Sentences (17)		Nicky put the ball in the net. Nigel is nine years old. Nancy has a sore neck. Nadia was making a noise Her name is Natasha. Animals that come out at night are nocturnal.	

Table 16 Treated /n/ Wordlist organised into word position, vowel environment, clusters and sentences

2.2.2.2.2.3 Additional Alveolar Wordlist

An additional alveolar list was also recorded at each of the six assessment time-points. The purpose of this wordlist was to sample /n/ within near-minimal pair sets, with more complex environments including a range of clusters. This wordlist contained 75 single words with 32 tokens of /n/, both singleton and clusters in WI, WM and WF positions (see Table 17). This wordlist also contained /n/ /ŋ/ minimal pairs that were targeted in therapy and comparable /t/, /s/, /m/, /ʃ/ and /k/ for analysis for example (know, toe, co, mole, snow). Some of the minimal pairs from this list were used in the treated wordlist. Words were elicited through an imitation task and were recorded using simultaneous ultrasound and audio data. Data from this wordlist were used for ultrasound analysis of minimal pairs.

Wordlist	Phonological Environment
knee know nap	WI /n/
tea toe tap	WI /t/
snow snack sneeze	WI /sn/
key co cap	WI /k/
meat mole magic	WI /m/
anteater aunty centre	WM /nt/
uncle blanket drinking	WM /ŋk/
dummy camel Emma	WM /m/
bumpy camping empire	WM /mp/
patting letter button	WM /t/
packing bucket record	WM /k/
messy castle bossy	WM /s/
brushing fashion seashells	WM /ʃ/
pansy disney	WM /nz/ vs. /zn/
pin sun fan	WF /n/
ping sung fang	WF /ŋ/
tonnes banned pinned	WF /n/ + suffix
tongues banged pinged	WF /ŋ/ + suffix
Tim sum ham	WF /m/
tent hunt ant	WF /nt/
pink skunk bank	WF /ŋk/
learn turn barn	WF /ɹn/
bench munch branch	WF /nt͡ʃ/
fence dance once	WF /ns/
plans pens balloons	WF /nz/

Table 17 Real-words recorded in additional alveolar wordlist

Non-words were used so that each target and comparison sound was also presented in isolation and in CV, VC and VCV syllables with the /a/ vowel (Table 18). The tokens containing /n/ were targeted in therapy.

Wordlist as presented	IPA (syllable position)
n an ana na	/n/ (isolation, VC, VCV, CV)
ng ang anga	/ŋ/ (isolation, VC, VCV)
t at ata ta	/t/ (isolation, VC, VCV, CV)
k ak aka ka	/k/ (isolation, VC, VCV, CV)
m am ama ma	/m/ (isolation, VC, VCV, CV)
s as asa sa	/s/ (isolation, VC, VCV, CV)
sh ash asha sha	/ʃ/ (isolation, VC, VCV, CV)

Table 18 Non-words recorded in additional alveolar wordlist

2.2.2.2.3 Intelligibility Measure

The Intelligibility in Context Scale (ICS, McLeod et al. 2012) was completed. See sub-section 2.1.7.3 for information on how the ICS was scored.

2.2.2.3 Post-Therapy Questionnaires

See method section (sub-section 2.1.8) for information on the post-therapy questionnaires. The post-therapy questionnaire for parents was completed in the post-VAM and post-UVBF assessment sessions. The post-therapy questionnaire for children was completed by Andrew and the SLT (the author) in the post-VAM and post-UVBF sessions and the three-month post-therapy questionnaire was completed in the maintenance session.

2.2.2.4 Therapy

2.2.2.4.1 Therapy Block One: VAM

Therapy followed an articulation therapy approach (Van Riper 1978) using the principles of motor learning (Maas et al. 2008). See method section (sub-section

2.1.9) for details regarding the basic therapy principles and feedback techniques. This block of therapy consisted of eight one-hour long therapy sessions using Speech Trainer 3D (Smarty Ears 2011) as an offline visual articulatory model (VAM) as a tool for training new articulations. Therapy block one targeted the production of /n/. Before targeting the production of /n/, the first and second therapy sessions focused on using Speech Trainer 3D to show Andrew the different parts of the vocal tract and to demonstrate and label sounds of English. During these sessions, auditory discrimination tasks using the iPad and visual discrimination tasks using still frames from the Speech Trainer 3D model were carried out, for example through games of *same/different* or *odd one out.* Andrew was asked to describe a sound by identifying if it was a *front* (alveolar) or *back* (velar) sound, a *mouth* (oral) or a *nose* (nasal) sound and a *loud* (voiced) or a *quiet* (voiceless) sound.

During each therapy session, Andrew's production practice was recorded in order for him to monitor his own progress and to identify his own errors using Speech Trainer 3D as an aid to demonstrate the correct and incorrect productions of his errors. For example, if he produced [ŋ] instead of [n], he would scroll through Speech Trainer 3D to find /ŋ/ to identify which sound he produced and would then describe how he had to change his tongue to produce an alveolar nasal stop by saying for example that he had to use the *front* of the tongue instead of the *back* of the tongue. Like in Preston et al. (2014) and Cleland et al. (2017c), therapy followed a hierarchy similar to that in Table 19. However, unlike the protocols in the literature, the current study did not follow the 80% accuracy level for moving up levels and therapy targets were more flexible during sessions, in line with Cleland et al. (2015c).

Level 0	CV or VC facilitative vowel
Level 1	CV
Level 1	VC
Level 2	CVC WI
Level 2	CVC WF
Level 3	Multisyllables
Level 4	Phrase repetition WI
Level 4	Phrase repetition WF
Level 5	Cloze (sentence completion)
Level 6	Clusters
Level 7	Complex sentences repetition and invention

Table 19 Hierarchy for Therapy Levels (Cleland et al. 2017c)

Table 20 provides information on which level within the hierarchy was worked on during each of the eight therapy sessions. Both auditory and visual discrimination was also worked on during therapy sessions one to four.

Therapy Session	Discrimination	0	1-CV	1-VC	2-WI	2-WF	3	4	5	6	7
1	X	X									
2	X	X									
3	X			X		X					
4	X										
5				X		X					
6			X		X						
7			X		X						
8					X		X				X

Table 20 Level Worked on in Each Therapy Session in Andrew's Therapy Block One. X indicates the level targeted in each session.

2.2.2.4.2 Therapy Block Two: UVBF

Pre-UVBF (assessment 4), Andrew scored 31% /n/ correct (see below for results).

The therapy target for therapy block two remained as /n/, to further improve PTCC.

Similar to therapy block one, the therapy approach used in block two was an articulation approach using the principles of motor learning. Speech Trainer 3D was replaced with ultrasound visual biofeedback.

The first therapy session focused on learning to associate the movement of the ultrasound video on the screen with the movement of Andrew's own tongue by demonstrating tongue shapes already in his inventory, such as /t/ vs. /k/ to demonstrate alveolar and velar placement. Ultrasound was used as visual articulatory model and ultrasound images were compared to those in Speech Trainer 3D through table top activities such as picture matching of static ultrasound and Speech Trainer 3D images (for example in games such as pairs or same/different) and matching of dynamic ultrasound to the dynamic videos in Speech Trainer 3D. The dynamic ultrasound videos were modelled by the SLT using a portable ultrasound machine.

Once Andrew was familiar with the orientation of the ultrasound images, we then moved on to showing Andrew figures of his tongue before and after therapy block one (Figure 11 and Figure 12) to show the improvement in his alveolar nasal in comparison to /t/ and /k/ as alveolar and velar references. On the images below, tongue tip is presented on the left side. It would be expected that /t/ and /n/ would have a similar tongue shape, in line with Gibbon et al. (2007). They found in their EPG study of normal adults that 99% of the time alveolar nasal and oral stops had the same spatial patterns, with the differences found in lateral contact. As ultrasound does not show lateral bracing in the midsagittal view, the images should show elevation of the tongue tip and lowering of the tongue dorsum for both /t/ and /n/, and lowering of the tongue tip and elevation of the tongue dorsum for /k/. Images in Figure 11 and Figure 12 are rotated by 20°, in line with Scobbie et al. (2011), which allows for comparison of analyses across different recordings.

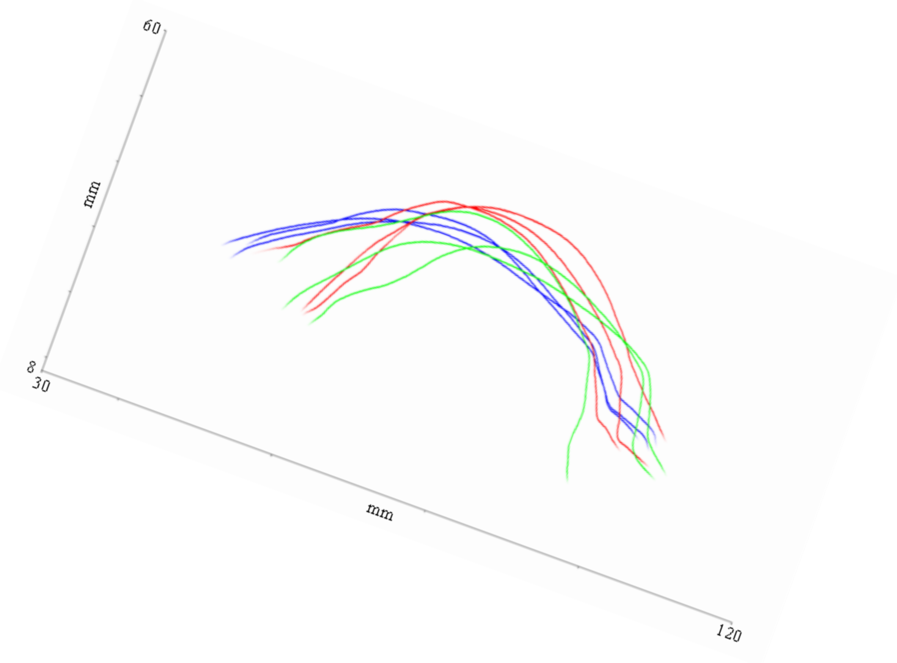


Figure 11 Figure shown to Andrew in therapy block two. /t/ /k/ /n/ pre-VAM in /i/ /o/ /a/ CV and CVC. /t/ - blue, /k/ - red, /n/ - green. Tongue tip on the left.

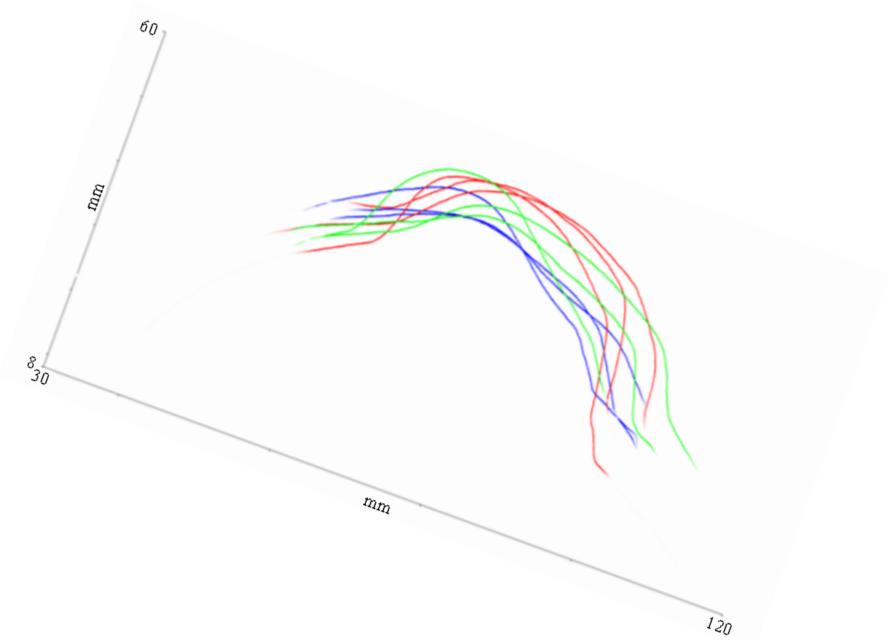


Figure 12 Figure shown to Andrew in therapy block two. /t/ /k/ /n/ post-VAM in /i/ /o/ /a/ CV and CVC. /t/ - blue, /k/ - red, /n/ - green. Tongue tip on the left.

It is evident from Figure 11 that /t/ and /n/ show a difference in tongue shape, with /n/ being closer to /k/, with lowering of the tongue tip and more movement in the tongue root zone (on the right of the image), therefore indicating incorrect productions of /n/. In Figure 12, /n/ has a raised tongue dorsum and tongue root, and is closer in shape to /k/ than /t/. Although there is also elevation in the tongue tip region that is not presented in Figure 11. When these figures were presented to him for explanation, Andrew was able to identify whether his productions of /n/ were closer to the tongue shape of /t/ or /k/, by comparing these figures to target tongue shapes on Speech Trainer 3D and using his own internal feedback.

Throughout therapy block two, it was suspected that Andrew had difficulty using the biofeedback element of ultrasound. Although static images of Andrew's production of /t/ and overlays were used, such as a cross for Andrew to keep his tongue below, Andrew found it particularly difficult to manipulate his tongue shape in order to achieve the target gesture for /n/. It was not until the latter sessions of therapy when an acetate was used as an overlay with a spline drawn on for an alveolar /t/ that Andrew began to use the biofeedback to achieve the target tongue shape correctly. It is possible that Andrew would have benefited from additional sessions to progress further, or would have benefited from an overlay of the whole tongue surface during sessions. Table 21 provides information on the level worked on in each therapy session during therapy block two.

Therapy Session	Discrimination	0	1-CV	1-VC	2-WI	2-WF	3	4	5	6	7
1	X	X	X								
2	X	X									
3	X	X	X								
4			X		X						
5					X			X			
6				X	X			X			
7											X
8											X

Table 21 Level Worked on in Each Therapy Session in Andrew's Therapy Block Two. X indicates level targeted in each therapy session.

2.2.3 Results

2.2.3.1 GOS.SP.ASS’98

The GOS.SP.ASS’98 (Sell et al. 1999) was repeated by the CLP SLT nine months after the maintenance session on the project. Results from the GOS.SP.ASS’98 showed that Andrew was producing [n] in SIWI position and substituting /n/ with [ŋ] in SFWF position. This, therefore, indicates improvement in the production of /n/ post-study. Pre-study, Andrew’s resonance was recorded as normal, with mild nasal turbulence. Post-study, Andrew’s resonance was recorded as mildly hypernasal, with mild nasal turbulence and mild inconsistent grimace. Table 22 summarises the GOS.SP.ASS results pre- and post-study for comparison, with green highlighted areas indicating sounds present in Andrew’s inventory post-therapy.

Four Months Pre-Study																			
	Labial					Alveolar						Post-Alveolar			Velar			Glottal	
	m	p	b	f	v	n	l	t	d	s	z	ʃ	ʧ	ʤ	ŋ	k	g	h	θð
SIWI						ŋ		t		s͌		ʃʲ		ʤ͌			g		
SFWF						ŋ		t͌		s		ʃˡ		ʤˡ			g↑		
Nine Months Post-Study																			
	Labial					Alveolar						Post-Alveolar			Velar			Glottal	
	m	p	b	f	v	n	l	t	d	s	z	ʃ	ʧ	ʤ	ŋ	k	g	h	θð
SIWI						n		t͌		s͌							g͌		
SFWF						ŋ		t		s									

Table 22 Andrew’s GOS.SP.ASS'98 Consonant Production Pre-and Post-Study Green shading indicates sounds present in Andrew’s inventory.

2.2.3.2 DEAP Phonology

Figure 13 shows Andrew’s PCC scores obtained from the DEAP Phonology subtest. No improvement was found between baseline and maintenance, with scores remaining stable within three percentage points (baseline = 87%; maintenance = 90%). Although the DEAP only has a small number of tokens of /n/ (N=8; elepha<u>n</u>t, trai<u>n</u>, ora<u>n</u>ge, quee<u>n</u>, s<u>n</u>ake, <u>kn</u>ife, va<u>n</u>, kitche<u>n</u>), it would be expected that the PCC score would increase. The grey shaded areas in Figure 13 denote periods of therapy.

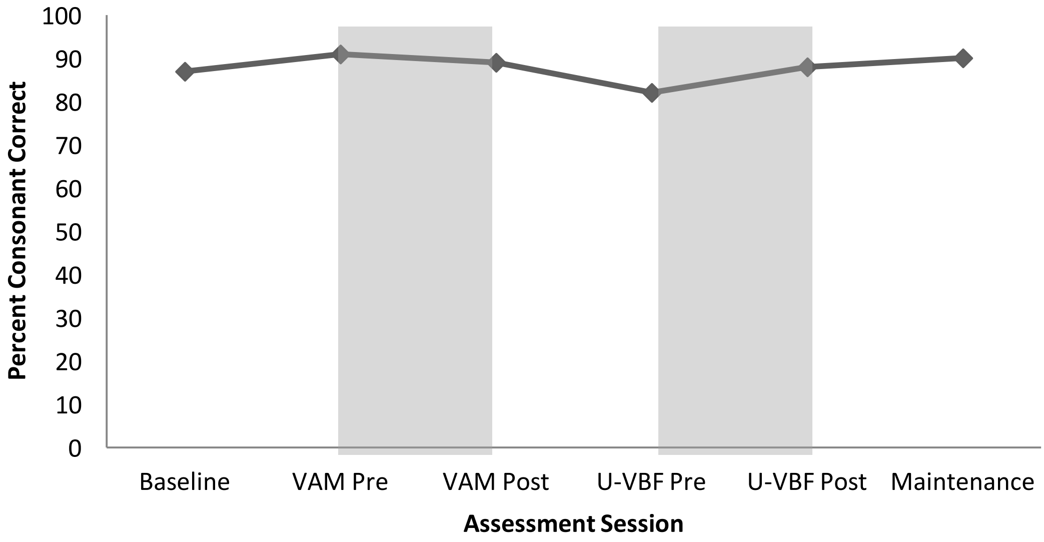


Figure 13 Andrew's DEAP Phonology subtest PCC scores from baseline to maintenance. Grey shading indicates period of intervention.

Table 23 provide details of the errors identified in the DEAP data, including typical patterns and atypical error patterns. These atypical errors, including cleft-type characteristics such as palatalisation, retraction and double articulations, make up 58% of Andrew’s errors in the DEAP overall, with retraction of /n/ to [ŋ] being the most consistent error.

NON-CLEFT PROCESSES	Specific Errors	Baseline	Pre-VAM	Post-VAM	Pre-UVBF	Post-UVBF	Maintenance
Fronting		6	5	6	8	8	5
Stopping		1					
Voicing errors		2	3		3	1	1
Cluster Reduction					1		
ATYPICAL PROCESSES							
Labialisation of sibilants				1			
Palatalisation	**/k g/ - [c ɟ]**				2	5	3
Backing	**/f/ - [θ]**			1			
	/s/ - [ʃ]	1			1	1	
	/s/ - [ç]				1	1	1
	/m/ - [ŋ]	1	1	1	1		1
	/n/ - [ɲ]						1
	/n/ - [ŋ]	4	3	2	5	5	3
	/t d/ - [k g]				1		
Double Articulations	**/n/ - [n͡ŋ]**			2	1	1	
	/k/ - [t͡k]			1	1		
Frication / Affrication	**/k/ - [ʧ]**	1					
	/k/ - [x]				2		
	/b/ - [ɸ]						1
Other	**Intrusive consonants**			1			
Total Typical Errors		9	8	6	11	9	6
Total Atypical Errors		7	4	9	15	13	10
TOTAL ERRORS		16	12	15	27	22	16

Table 23 Andrew's DEAP Phonology Error Pattern Analysis, separating non-cleft processes and atypical processes more commonly associated with CP

2.2.3.3 Treated /n/

As there were a small number of /n/ tokens, target-specific wordlists were also recorded. Figure 14 shows Andrew's PTCC scores obtained from the treated wordlist from post-VAM through to maintenance, transcribed by the treating SLT (the author). As reported above, there were no baseline or pre-VAM recordings for the treated wordlist as this list was devised during therapy block one. In the Post-VAM assessment Andrew had a PTCC score of 24%, which decreased to 5% between post-VAM and pre-UVBF (five-week gap with no therapy). Post-UVBF, Andrew's score increased to 26%, with a further increase to 41% during the three-month maintenance period. As there was no therapy between post-UVBF and maintenance, this would suggest continuing generalisation in the maintenance period. In accordance with Preston et al. (2014) an increase from 24% post-VAM to 41% in maintenance would suggest that Andrew's increase in PTCC in the treated wordlist is not clinically significant, however there is an obvious upward trend.

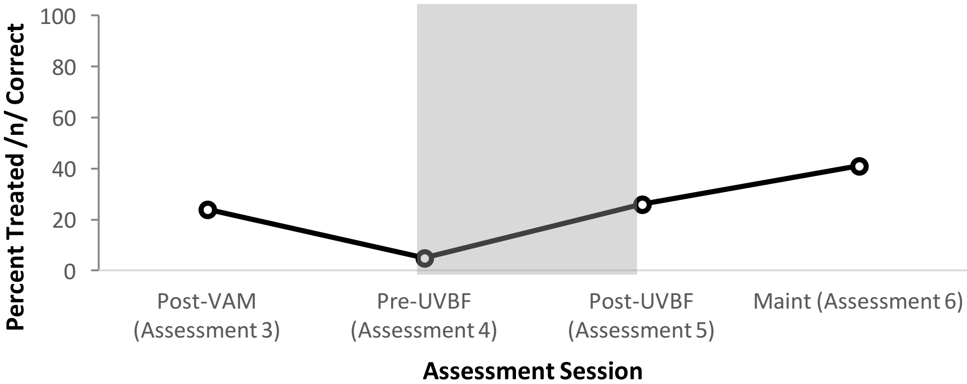


Figure 14 Andrew's Treated /n/ PTCC scores across four assessment time-points (post-VAM to Maintenance). Grey shading indicates period of intervention.

2.2.3.4 Untreated /n/ Wordlist

Figure 15 shows Andrew's PTCC scores obtained from the 36 single words and six sentences across all six assessment time-points, transcribed by the treating SLT. When looking at the PTCC scores from single words, at baseline Andrew achieved a PTCC score of 5% which remained relatively stable in the pre-VAM assessment (8%). In the Post-VAM assessment Andrew's PTCC had increased to 21%. Scores increased between the post-VAM and the pre-UVBF assessment (31%). After the second block of therapy this had decreased to 16% in the Post-UVBF session, which increased by five percentage points to 21% in the maintenance session. The biggest improvement was found after the VAM comparison, with improvement overall from baseline to maintenance. Preston et al. (2014) suggests that 20% improvement is clinically significant. The overall comparison of Andrew's PTCC scores from 5% at baseline to 21% in maintenance would suggest that results are not clinically significant at single-word level. However, it does show that Andrew was able to acquire /n/ at single word level using Speech Trainer 3D.

Twenty-five tokens of /n/ were scored in the sentences. At baseline, Andrew had a PTCC score of 0%, which increased to 4% pre-VAM. Post-VAM the PTCC score increased to 20%, which decreased to 16% pre-UVBF. Post-UVBF the PTCC score in sentences had increased to 32%, which then decreased to 16% in the maintenance session. An increase from 0% to 16% again shows a clinically non-significant improvement in Andrew's production of /n/ in sentences overall (Preston et al. 2014). Unlike the PTCC scores at single-word level, Andrews PTCC scores at sentence level increased in the UVBF comparison.

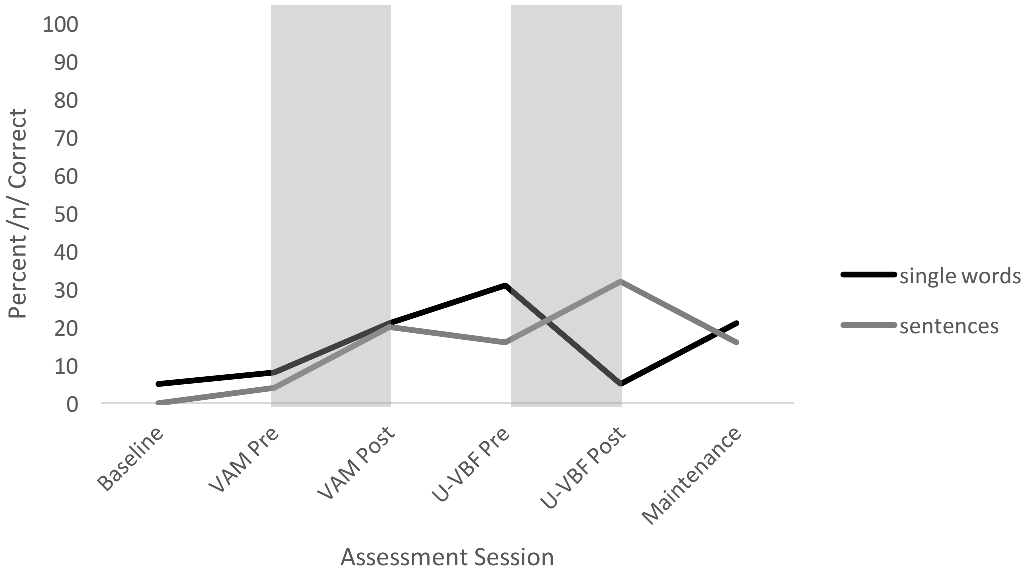


Figure 15 Andrrw's PTCC Scores from the Untreated /n/ wordlist (single words and sentences). Grey shading indicates period of intervention.

2.2.3.4.1 Phonological Environment Analysis

Overall, 16 out of 39 tokens of /n/ were produced correctly in at least one session, with 23 tokens being incorrect 100% of the time. Correct productions occurred most commonly in word initial position across all six assessments. Assessments four and six (pre-UVBF and maintenance) had the highest number of correct productions in WI position (6/12), Pre-UVBF had the highest number of correct productions in all word positions (13/39). Table 24 provides an error pattern analysis for all tokens of /n/ in WI, WM and WF position.

	Baseline	Pre-VAM	Post-VAM	Pre-UVBF	Post-UVBF	Maintenance
WI singleton	[n] (1) [ŋ] (10) [dɹ] (1)	[n] (2) [ŋ] (10)	[n] (4) [ŋ] (5) [n͡ŋ] (3)	[n] (6) [ŋ] (6)	[n] (1) [ŋ] (11)	[n] (6) [ŋ] (1) [n͡ŋ] (5)
WI /sn/	[ŋ] (1)	[ŋ] (1)	[ŋ] (1)	[n] (1)	[ŋ] (1)	[ŋ] (1)
WM Singleton	[ŋ] (12)	[ŋ] (12)	[ŋ] (12)	[n] (2) [ŋ] (10)	[ŋ] (12)	[n] (1) [ŋ] (11)
WM /nj/	[ŋ] (1)	[ŋ] (1)	[ŋ] (1)	[ŋ] (1)	[ɲ] (1)	[ŋ] (1)
WF singleton	[n] (1) [ŋ] (10)	[n] (1) [ŋ] (10)	[n] (4) [ŋ] (7)	[n] (3) [ŋ] (8)	[n] (1) [ŋ] (10)	[n] (1) [ŋ] (6) [n͡ŋ] (4)
WF /ɹn/	[ŋ] (1)	[ŋ] (1)	[ŋ] (1)	[ŋ] (1)	[n͡ŋ] (1)	[ŋ] (1)
WF /n/ + suffix	[ŋ] (1)	[ŋ] (1)	[ŋ] (1)	[n] (1)	[ŋ] (1)	[ŋ] (1)

Table 24 Andrew's Error Pattern Analysis for /n/ tokens in single words *N.B brackets indicate the number of occurrences

Table 25 shows the words in which Andrew produced /n/ correctly within each of the six assessment sessions. The word "kneeling" was produced correctly pre-therapy and remained correct in 3/4 of the remaining assessments (minus post-UVBF). "nibbling" was also produced correctly pre-therapy and remained correct in 3/4 of the remaining sessions (minus pre-UVBF).

	Word Produced Correctly	Baseline	Pre-VAM	Post-VAM	Pre-UVBF	Post-UVBF	Maintenance
WI singleton	Kneeling	X	X	X	X		X
	Nibbling		X	X		X	X
	Knitting			X	X		X
	Neeps			X	X		X
	Nose				X		X
	Necklace				X		X
	Nuts				X		
WI /sn/	Snowman				X		
WM singleton	Animals				X		
	Brownie				X		X
WM /nj/							
WF singleton	Medicine	X		X			
	Garden		X	X			
	Violin			X	X		X
	Green				X	X	
	Curtain			X	X		
WF /ɹn/							
WF /n/ + suffix	Onions				X		

Table 25 Words in which Andrew produced [n] correctly (X indicates correct production)

Figure 16 shows an analysis of WI singleton /n/. The correct production of WI /n/ increases from baseline (N=1) to maintenance (N=6). Whilst the number of correct [n] increases steadily across time, there is an obvious decrease post-UVBF (N=1). It is clear that the most common substitution for WI /n/ is [ŋ], which decreases from baseline (N=10) to maintenance (N=1), with only two sessions with double articulations (post-VAM N=3; Maintenance N=5). WI has the most correct productions across all three word positions (WI, WM and WF).

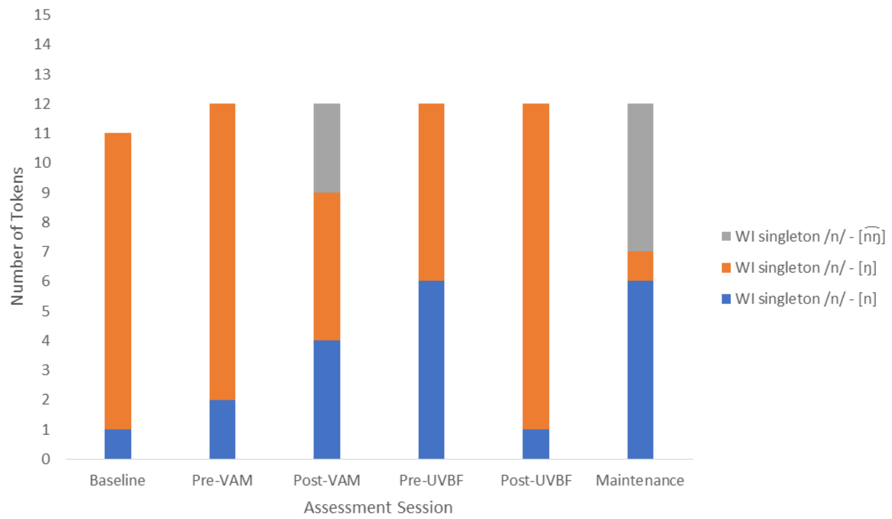


Figure 16 Andrew's Untreated /n/ Error Pattern Analysis (WI Position)

Unlike WI position, there are fewer correct productions of /n/ in WM position, as shown in Figure 17, with mostly all WM tokens perceived as [ŋ]. Only two sessions have correct productions (pre-UVBF N=2; maintenance N=1). No double articulations were transcribed word medially.

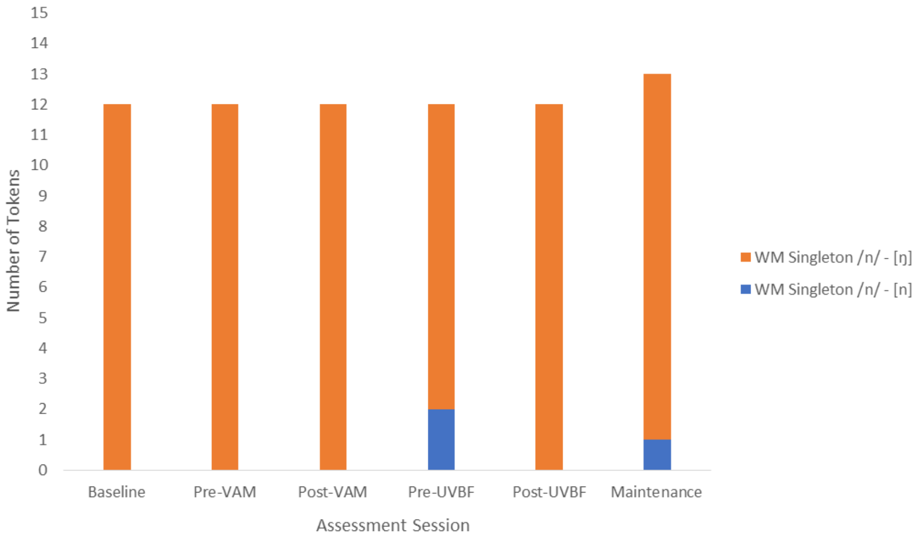


Figure 17 Andrew's Untreated /n/ Error Pattern Analysis (WM Position)

Interestingly, word finally, there is an increase from pre-VAM (N=1) to post-VAM (N=4), however this decreases post-UVBF and maintenance (N=1) (see Figure 18). Double articulations were only detected in the maintenance session (N=4), with [ŋ] being the most common substitution.

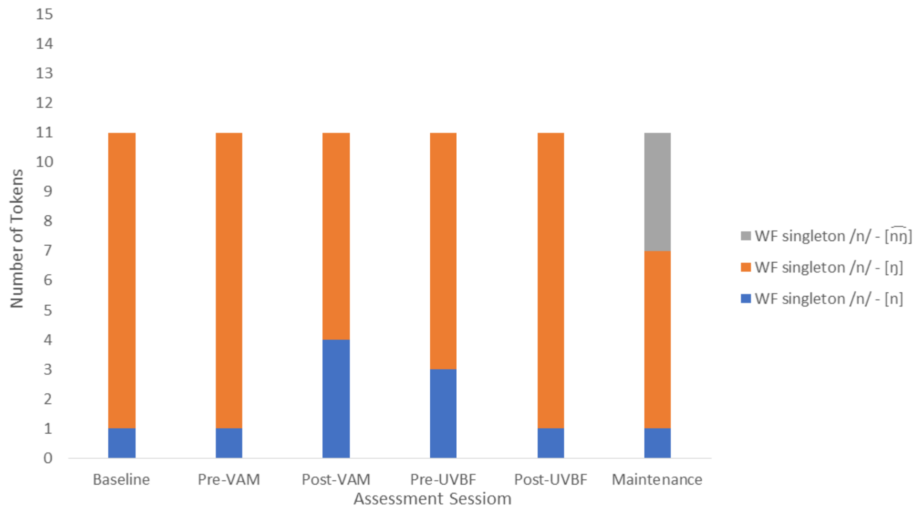


Figure 18 Andrew's Untreated /n/ Error Pattern Analysis (WF Position)

2.2.3.4.2 Intra-rater reliability of Untreated /n/

The researcher transcribed all 36 single words twice, with a three-year interval between transcription one and transcription two. Results showed agreement across the two time-points (mean=75.5%, range=63%-87%). Figure 19 shows percentage of agreement across all six assessment time-points.

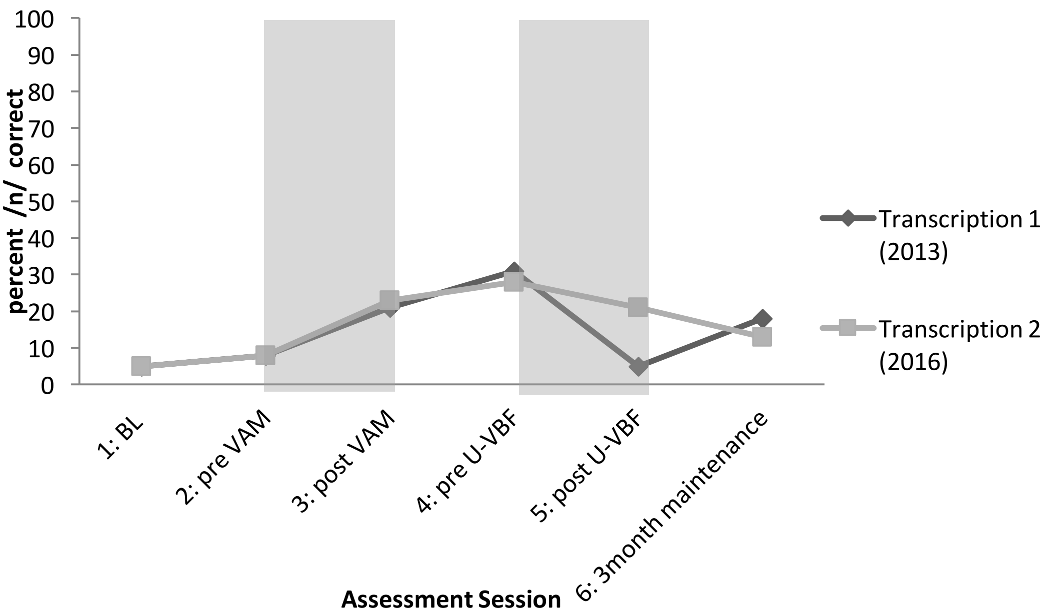


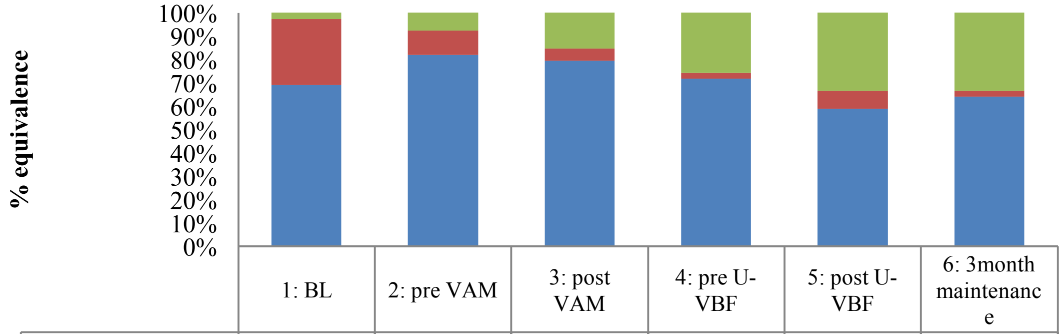
Figure 19 Andrew's Intra-Rater Reliability. Grey shaded areas indicate periods of intervention.

Statistical analysis showed that the highest agreement was found in the baseline session (Cohen's Kappa=1 "perfect agreement") with the lowest agreement found in the Post-UVBF session (Cohen's Kappa=0.128 "poor agreement"). Table 26 shows Cohen's Kappa results for all six assessment sessions.

Session	no. observed agreements	no. expected by chance	Kappa Score	SE of Kappa	95% CI	Strength
baseline	39 (100%)	35.2 (90.27%)	1	0	1.000 to 1.000	perfect
Pre-VAM	37 (94.87%)	33.5 (85.80%)	0.639	0.236	0.176 to 1.000	good
Post-VAM	36 (92.31%)	25.7 (65.88%)	0.775	0.124	0.532 to 1.000	good
Pre-VBF	30 (76.92%)	22.8 (58.38%)	0.445	0.156	0.139 to 0.752	moderate
Post-VBF	31 (79.49%)	29.8 (76.46%)	0.128	0.163	-0.191 to 0.448	poor
maintenance	37 (94.87%)	28.8 (73.83%)	0.804	0.132	0.545 to 1.000	very good

Table 26 Andrew: Intra-Rater Cohen's Kappa Scores

Equivalence scores for transcriptions over the two time-points (2013 and 2016) (see sub-section 2.1.10.1) were calculated using a three-point scale (2=same; 1=almost equivalent; 0=different). Figure 20 shows equivalence scores for each assessment time-point and percentages of each equivalence score.



	1: BL	2: pre VAM	3: post VAM	4: pre U-VBF	5: post U-VBF	6: 3month maintenance
0: different	1	3	6	10	13	13
1: almost equivalent	11	4	2	1	3	1
2: same	27	32	31	28	23	25

Figure 20 Andrew: Intra-Rater Equivalence Scores

2.2.3.4.3 Inter-rater reliability of untreated /n/

All 36 single words were transcribed by a further two phoneticians, who were blinded to the assessment time-point. Inter-rater reliability results show that, when compared to the author's transcriptions, all three transcribers agreed the majority of the time (mean=72% range=59%-80%). Statistical analysis showed that the highest agreement across transcribers was found in the Maintenance session (Fleiss' Kappa = .6538 "intermediate to good") with all three transcribers agreeing on 31/39 tokens. The lowest agreement was found in the pre-VAM session (Fleiss' Kappa = .0969), despite all three transcribers agreeing on 30/39 tokens. Figure 21 shows PTCC scores derived from all three transcribers. Equivalence scores were not calculated for inter-rater reliability due to differences in transcription techniques (i.e. broad vs. narrow transcriptions).

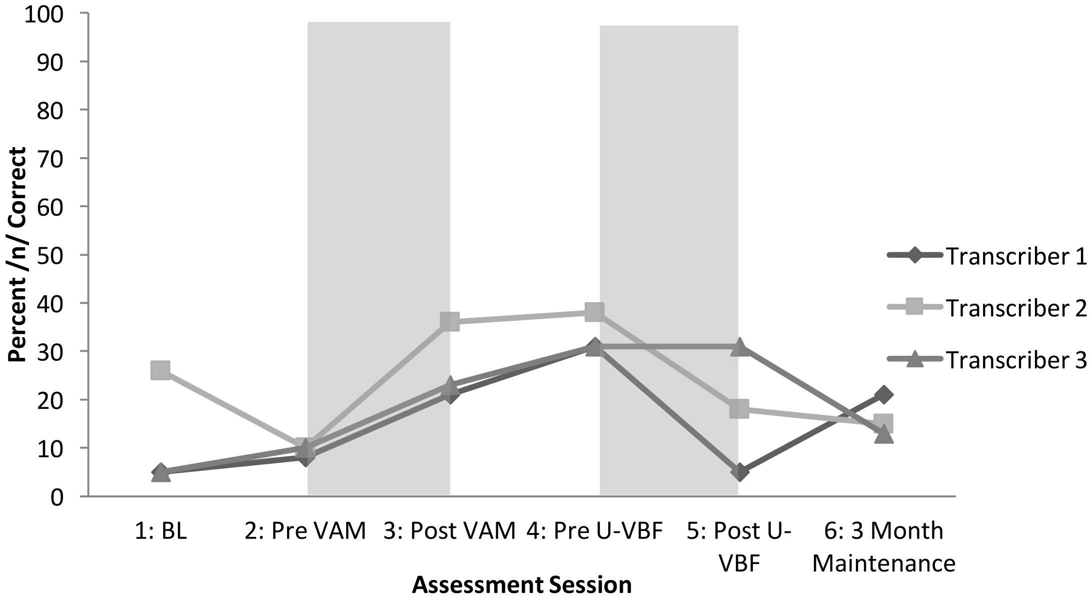


Figure 21 Andrew's PTCC scores for Inter-Rater Reliability (Transcriber 1 = treating clinician). Grey areas indicate periods of intervention.

2.2.3.5 Additional Alveolar Wordlist

Results from the additional alveolar wordlist (recorded to sample /n/ within near-minimal pair sets, with more complex environments including a range of clusters), are separated from the Untreated wordlist as these were not included in inter-rater or intra-rater reliability and also due to some of the words in this list being treated in therapy. Figure 22 shows Andrew's PTCC scores obtained from the additional alveolar wordlist across six assessment time-points, transcribed by the treating SLT (the author). At baseline, he achieved a PTCC score of 11% which remained relatively stable in the pre-VAM assessment (8%). In the Post-VAM assessment Andrew's PTCC had increased to 16%. Scores decreased between the post-VAM (16%) and the pre-UVBF assessment (5%). After the second block of therapy this had increased to 13% in the Post-UVBF session, with a further increase to 21% in the maintenance session. Overall, there was an increase from 11% to 21%, suggesting a clinically non-significant improvement in /n/ in the additional alveolar wordlist.

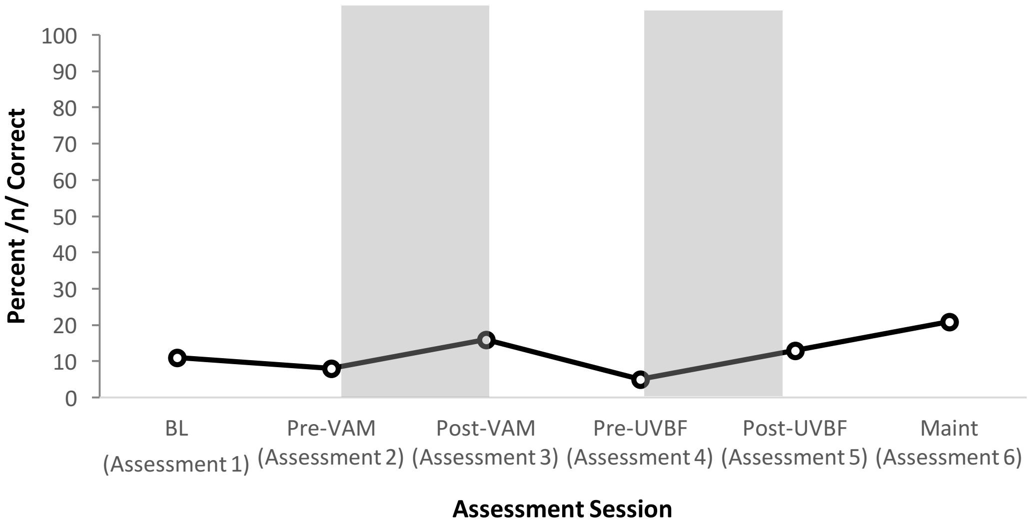


Figure 22 Additional Alveolar Wordlist: PTCC scores across all six assessment time-points. Grey shading indicates period of intervention.

2.2.3.6 Questionnaires

2.2.3.6.1 Intelligibility in Context Scale

The ICS showed stability across all three comparisons. At each assessment time-point, Andrew's parents reported that all listeners "usually" understood Andrew. There was no difference between any of the listeners (i.e. parents or strangers).

2.2.3.6.2 Therapy Outcome Questionnaire for Parent: Parental Responses

Post-VAM Therapy: Andrew's Mother reported that his speech had *greatly improved* since enrolling on the project. She reported that after therapy block one, his awareness of speech sounds had *greatly improved,* and his ability to articulate the target sounds had *moderately improved.* It was reported that post-VAM, he was able to use /n/ *some of the time* in conversation. Compared to his siblings, his Mother rated his speech as *slightly better* since attending therapy. When asked "please comment on whether or not you think using Speech Trainer 3D has made it easier for your child to achieve his speech therapy goals", his Mother reported that she felt Speech Trainer 3D to be a useful tool and that it allowed Andrew to view a correct representation of /n/. She reported that she felt it was a very usable tool and that it was relevant for children as they all love iPads. She also reported that using Speech trainer 3D made her more aware of the difference in pronunciation of /n/ and /ŋ/, particularly when he is producing his sounds in single words.

Post-UVBF Therapy: Andrew's Mother reported that his speech and specifically his target /n/ had *greatly improved*. His Mother reported that after therapy block two, his awareness of speech sounds and his ability to articulate the target sounds also *greatly improved*. It was reported that post-UVBF, he was able to use [n] *some of the time* in conversation, which had not improved since post-VAM. Compared to his siblings, his Mother rated his speech as *slightly worse*. When asked "please comment on whether or not you think using ultrasound has made it easier for your child to achieve his speech therapy goals", his Mother reported that she felt being able to see his own tongue in real time helped Andrew to modify his own productions.

VAM vs. UVBF: Andrew's Mother also commented that both the iPad and ultrasound had their advantages. She felt that the iPad would be useful for local SLTs who would not have access to the ultrasound facilities. She noted definite improvement in Andrew's speech and that she felt taking part in the study had been very beneficial. As ethical approval did not cover the use of direct quotes, actual parental responses are not included as an appendix.

2.2.3.6.3 Therapy Outcome Questionnaire for Children: Participant Responses

Post-VAM Therapy: Andrew reported that he felt using the iPad was very interesting and that he enjoyed using the iPad because he was able to see what his tongue should be doing. He reported that the worst/hardest bit about using the iPad was when he was first introduced to the visual articulatory model and asked to discuss differences between /n/ and /ŋ/. He felt that using the iPad helped his speaking. He said it helped him because he was able to see all of the different sounds and their tongue shape, and it was good because iPads are suited to children. When asked if the sessions were "too short, just right or too long" he reported that they were just right. These sessions lasted around one-hour.

Post-UVBF Therapy: Andrew reported that he thought looking at his tongue was really useful to see the correct shape of his own tongue and that it helped him a lot. He reported that he enjoyed seeing his own tongue moving during the ultrasound sessions and that it helped with his speech. He said the worst bit about the ultrasound was getting the headset on and the hardest bit was trying to keep his tongue dorsum lowered. When asked if the sessions were "too short, just right or too long" he reported that they were "just right". Each session lasted an hour, with 30 minutes spent using ultrasound wearing the headset and 30 minutes doing table-top activities. Table 27 shows Andrew's responses to questions regarding intelligibility with a range of listeners for both the post-VAM and post-UVBF sessions.

	Post-VAM	**Post-UVBF**
How often do you think your parents understand you when you speak?	Almost Always	Almost Always
How often do you think your brothers or sisters understand you when you speak?	Rarely	Almost Always
How often do you think your teacher at school understands you when you speak?	Always	Always
How often do you think your friends understand you when you speak?	Always	Always
When you talk to new people, how often do they understand you when you speak?	Almost Always	Almost Always

Table 27 Andrew's Responses to Intelligibility Questions in Post-Therapy Questionnaires

Maintenance: Andrew reported that he preferred using the iPad to ultrasound. When asked why he preferred the iPad, he reported that the ultrasound was uncomfortable. When asked which tool was easier to use he chose ultrasound because he was able to see his own tongue. As ethical approval did not cover the use of direct quotes, actual responses from Andrew are not included as an appendix.

2.2.4 Clinical Discussion

The following section will reflect on the use of Speech Trainer 3D and ultrasound visual biofeedback for Andrew, with reference to the objective PTCC scores derived from transcriptions, ICS scores and questionnaire responses. It will also consider the researcher's perspective throughout.

2.2.4.1 Therapy Outcomes

Andrew was referred to the project by the CLP specialist SLT who reported that Andrew was retracting /n/ to velar nasals with the possibility of double articulations.

Previous studies of VAMs showed little to no improvement in speech outcomes (Gotto 2004; Albert 2005) or improvement in errors that were not treated for Andrew or Craig (Fagel and Madany 2008). Similarly, when VAMs have been used for second language learning, there has been no advantage found for using VAMS to teach lingual movements such as velar or uvular plosives (Massaro et al. 2008). Cleland and Scobbie (in press) argue that VAMs alone are not the "key ingredient" for learning new articulations. Therefore, it was hypothesised that Andrew's PTCC scores would remain stable during therapy block one and would increase after therapy block two using ultrasound visual biofeedback.

PTCC scores derived from phonetic transcription of the untreated wordlist showed an increase in scores from initial baseline to maintenance, three months after therapy ceased. For Andrew, this improvement was modest, rising from 3% PTCC at baseline to only 17% PTCC at maintenance. This is unlikely to represent a clinically significant improvement in Andrew's production of /n/ suggesting that neither therapy was particularly effective. An increase in PTCC scores was found in therapy block one using Speech Trainer 3D (pre-VAM: 6%; post-VAM: 20%), with a decrease in PTCC scores identified in therapy block two using ultrasound (pre-UVBF: 23%; post-UVBF: 16%). PTCC scores from the treated wordlist increased to 41% in maintenance; however, the additional alveolar list suggests very modest improvement in Andrew's production of /n/ in a range of singleton and consonant cluster productions in all word positions. PCC scores in the DEAP Phonology subtest also remained stable across all six sessions.

Despite the low increase in PTCC scores, Andrew's parents reported they felt his speech has 'greatly improved' during the course of therapy, both using Speech trainer 3D and using Ultrasound visual biofeedback, and was using his target sound in conversation 'some of the time'. Andrew reported that using ultrasound was easier as he was able to see his own tongue, but that he preferred using Speech Trainer 3D as wearing the headset with the ultrasound was uncomfortable.

As it was hypothesised that PTCC scores would remain stable during therapy block one and increase after therapy block two, it was surprising when Andrew was able to achieve his target in therapy block one. However, his productions were difficult to transcribe perceptually (because of ambiguity of place of articulation and resonance difficulties such as VP friction) and due to a lack of biofeedback during therapy it was difficult to determine whether Andrew was producing /n/ correctly or if he had undifferentiated lingual gestures (Gibbon 1999) or double articulations (Gibbon 2004). Andrew was able to label sounds, for example /n/ is a *front, loud, nose sound* and was able to describe and draw what he felt his own tongue was doing during his own productions. He described his productions as "going up at the front and back and that it goes out at the sides", suggesting possible double articulations or temporal difficulties. Andrew also commented on his awareness of his surgical scar during therapy block one using Speech trainer 3D. As Andrew became more aware of his own productions and his errors, his frustration levels also increased. Andrew's Mother reported that using the visual models helped her understand the phonetic descriptions she had previously heard Andrew's SLT's using, e.g. 'front/back' and also made her more aware of Andrew's speech errors perceptually. As UTI images were not viewed by the SLT or participants during baseline and pre-VAM and the ultrasound data was not analysed until post-VAM, it was felt necessary that further investigation using UTI was required for both assessment and therapy in block two.

Further to progress in therapy block one, Andrew's PTCC scores decreased during the ultrasound therapy block when it was hypothesised scores would increase. During the course of treatment using ultrasound biofeedback, Andrew became increasingly more frustrated when he was able to monitor his own tongue movements and see that he was not able to achieve his target tongue shape. As the treating clinician, it was at times difficult to provide KP feedback, due to the poor image quality. This may have been the result of the tongue tip image being partially lost due to a large mandible shadow or because the headset was difficult to fit, due to Andrew's facial asymmetry. This made treating an alveolar sound particularly difficult as the alveolar region was not always imageable. Andrew started to become more consistent in his productions toward the end of the block of therapy. It was felt that he would have benefited from further sessions or from the use of whole tongue overlays during the course of therapy.

2.2.4.2 Difficulties with Phonetic Transcription

Due to the nature of Andrew's speech errors and the possible double articulations, transcribing the data was particularly challenging. When the data was first collected in 2013, the treating SLT (the author) did not have wide experience in transcribing disordered speech data associated with CP, with only one module of training three years previously during the undergraduate SLT course. The two additional transcribers also had limited experience transcribing data associated with CP. While it would have been beneficial to have CP specialists transcribing the data, this was not feasible within the current study. Despite lack of training or expertise in transcribing data from children with CP, inter-rater and intra-rater reliability showed relatively good agreement across transcriptions, with the exception of the pre-VAM (inter-rater) and post-UVBF (intra-rater) sessions. When transcribing the intra-rater reliability data, the treating clinician was more experienced in treating and transcribing the speech of children with primary SSDs and secondary SSDs associated with CLP, which made the researcher more confident in transcription abilities.

2.2.4.3 Evaluation of Therapy Tools

The results from PTCC scores and from questionnaires suggest improvement in Andrew's production of /n/ overall from baseline to maintenance with more progress made in therapy block one using Speech Trainer 3D than with ultrasound visual biofeedback. When using Speech Trainer 3D, Andrew had the references for other areas of the vocal tract. The SLT was able to instruct Andrew using the model to identify the passive articulators (i.e. alveolar ridge). When using ultrasound, the tongue tip/alveolar region was often missing. Andrew had a small space under the chin, causing a large mandible shadow on the ultrasound images. Andrew also has hemifacial macrosomia and unilateral microtia. Due to the structural difficulties, such as facial asymmetry the probe-stabilising headset was not able to sit straight on Andrew's head. This often created artefacts on the ultrasound image or a skewed image which was not in the midsagittal plane. As the tongue tip image was missing, this made it difficult to provide the correct feedback during therapy sessions. An overlay was used to provide a reference for Andrew to keep the back of his tongue lower than the cross. However, due to the quality of the image it was not possible to add on a palate trace. Therefore, unlike with Speech Trainer 3D there were no references to the passive articulators.

Andrew was given a reference of his own production of [t] as a model for the tongue shape of an alveolar sound. This reference was a static image. As technology has advanced, references now come in the form of dynamic videos which may have been more beneficial for Andrew.

2.2.4.4 Evaluation of Speech Materials

During the evaluation of the speech materials, there were obvious strengths and gaps noted in the untreated, treated and additional alveolar wordlists. Firstly, considering the untreated /n/ wordlist, there were 36 words containing 39 tokens of /n/. A range of vowels were used where possible; however, it was felt that this wordlist should have included words with more complex structures, i.e. multisyllabic words and more clusters. Nevertheless, this was unproblematic as the additional alveolar wordlist, recorded separately to the untreated /n/ wordlist, included multiple clusters and also minimal pairs. When combining both the untreated /n/ and the additional alveolar wordlist, a total of 71 tokens of /n/ were measured during each of the six assessment time-points.

It may have been useful to prepare the treated wordlists in advance in order to record these with ultrasound at baseline. As data was not recorded pre-therapy, a pre-/post-therapy comparison of treated tokens to assess acquisition and retention was not possible. However, by tailoring this wordlist each week during therapy block one, it ensured that the materials were specific to Andrew's needs during therapy, as therapy progressed through tasks of increasing complexity. The treated wordlist could be improved by adding in tokens of inter-vocalic /n/.

2.3 Craig

2.3.1 Background Information

2.3.1.1 General clinical profile

Craig (pseudonym) is a 6;2 year-old male. He was born at 36 weeks gestation and has a hypoplastic hand. At nine months of age, a very posterior cleft was identified, which was thought at the time to be no more than a bifid uvula and submucous cleft. At the time of his first MDT assessment, there was a degree of asymmetry, with the left side of his palate being shorter than the right. It was also reported from his initial assessment with the MDT that he had a narrow hard palate. As there were no airway problems at this stage and there was no nasal regurgitation during eating and swallowing, surgery was thought not necessary at this early stage, however was repaired at a later age of 2;6 years. He then required secondary surgery at age 5;10.

Craig has received significant involvement from the multidisciplinary cleft team, including speech and language therapists, and from various other services such as occupational therapy, respiratory clinic and community child health. At the time of referral to the project, the specialist CLP SLT reported that he had a limited range of high pressure consonants and presented with cleft-type characteristics such as retraction to glottal placement and suspected double articulations. He had mild hypernasality with accompanying mild nasal turbulence. His CLP SLT requested velar plosives as the therapy target on the project (see below for rationale).

2.3.1.2 Chronology of Craig's MDT interventions and diagnosis

Craig has been known to the CLP services from age 0;9. Table 28 summarises the input he has received between birth and the time of referral to the current research project at age 6;2. The key information from Craig's chronology is summarised below. Information is taken from case-notes from the CLP Specialist SLT and is organized into SLT input, MDT assessment and surgery.

Age (years;months)	SLT Input	MDT Input	Surgery
0;9		MDT Ax – SMCP diagnosed	
1;0	CLP SLT Ax		
1;2		Consultant report	
1;4	Specialist SLT – Eating, Drinking and Swallowing (EDS) Ax		
1;5		Referral for videofluoroscopy (VF)	
1;6	Specialist SLT – EDS Ax (no VF required)		
1;7		Respiratory Ax	
1;9	CLP Specialist Ax – referred to community SLT		
2;1	Community SLT report		
2;4		Consultant Report	
2;5		Consultant Report	
2;6			Primary Surgery to repair SMCP
2;8		Consultant Review	
2;10	Community SLT Report		
3;6	CLP SLT Ax		
3;1-4;2	Community SLT Input		
4;2	CLP SLT Ax		
4;4		Ax with Community Child Health (CCH)	
4;9		MDT Ax	
4;10		Hearing Ax	
4;11		Occupational Therapy Ax	
5;2		Ear, Nose and Throat Ax – referred for tonsillectomy	
5;3	Community SLT Report – ongoing tx		
5;4		CCH Review	
5;10			Secondary Surgery – Hynes Pharyngoplasty
6;0		MDT Ax	
6;2	CLP SLT – referred to Research Project		

Table 28 Summary of Craig's input from birth to referral to the current project

2.3.1.3 Specialist and Community SLT Input: Assessment and Therapy

At his initial assessment with the CLP SLT at age 1;0, it was reported that Craig vocalised on occasion, which consisted of undifferentiated vowels, and he made some grunts and squeals. At the time of his initial assessment, he did not attempt production of any oral consonants.

Craig was reassessed by the CLP team at age 1;9. During this assessment, language difficulties were detected and Craig was referred to the community SLT. At this stage, it was not possible to assess the impact of his SMC on his speech development due to limited language and speech sound inventory. Craig continued to receive support from the community and specialist CLP SLT. By age 4;2 it was reported that all consonants were realised as nasals with occasional palatal friction. However, he had made limited progress with a large amount of input and despite having surgery at age 2;6.

Craig continued to receive SLT input, working on his phonological awareness skills and production of oral pressure consonants. By age 5;3, the community SLT reported that his language skills had improved. He was beginning to produce [f] in all word positions but was not yet using lingual fricatives. /d/ was still being realised as nasal [n] and it was effortful for Craig to produce other lingual consonants. As a result of his continuing difficulties with speech production, Craig went on to receive secondary surgery at age 5;10. Table 29 shows results of Craig's consonant production from the GOS.SP.ASS'98 (Sell et al. 1999), which was completed by the CLP SLT at his MDT assessment post-surgery at age 6;0, two months prior to him starting on the research project. Correct consonants are highlighted in green.

	Labial					**Alveolar**						**Post-Alveolar**			**Velar**			**Glottal**	
	m	**p**	**b**	**f**	**v**	**n**	**l**	**t**	**d**	**s**	**z**	**ʃ**	**ʧ**	**ʤ**	**ŋ**	**k**	**g**	**h**	**θð**
SIWI		ʔ͡p	b͌	(f)͡ʔ	ɔ			ʔ	n	n͡ʔ	n	n͡ʔ	ʔ	n		ʔ	n		ð
SFWF		p͡ʔ	b	f͌	v			ʔ	n	s↓		s̪↓	(t)͡ʔ	ħ	n	ʔ	n		

Table 29 Craig's GOS.SP.ASS'98: Craig's consonant production pre-study. Green shading indicates sounds present in Craig's speech inventory.

From the GOS.SP.ASS'98, it is evident that Craig had a very limited phonetic inventory. As there was no lingual approximation for /k/ or /t/, velar plosives were the targets for therapy block one, with the addition of alveolar /t/ in therapy block two of the current study. The GOS.SP.ASS data here shows that there is a collapse in contrast between /t/ and /k/ which are phonetically transcribed as [ʔ] and /d/ and /g/

which are phonetically transcribed as [n]. During the course of the project, Craig also received a block of therapy from the CLP specialist SLT targeting his production of /s/.

2.3.2 Method

In the current study, Craig received six assessment sessions and two blocks of therapy, each with eight one-hour therapy sessions. See section 2.1 for general procedure and recording set-up. Details of therapy will be described below.

2.3.2.1 Language and Non-Verbal Measures

Craig attended two one-hour long baseline sessions prior to therapy block one. Language measures (see sub-section 2.1.6) were completed during these two sessions. Craig's receptive vocabulary measured in the normal range, with a standard score of 90 in the BPVS-III (Dunn et al. 2009). His language score was in the normal range, with a standard score of 93 in the CELF-4UK core language subtests (Semel, et al. 2006). Table 30 shows a breakdown of individual subtests from the CELF-4UK. Craig's non-verbal IQ was within the 75th percentile, also within the normal range for his age (Raven's Coloured Progressive Matrices (Raven et al. 1998).

Subtest	Raw Score	Scaled Score	Scaled Score Points +/-	Confidence Interval 90% level			Percentile Rank	Percentile Rank confidence interval		
Concepts and Following directions	36	11	1	10	**to**	13	63	50	**to**	75
Word Structure	19	8	2	6	**to**	10	25	9	**to**	25
Recalling Sentences	28	7	1	6	**to**	8	16	9	**to**	25
Formulated Sentences	23	9	2	7	**to**	11	37	16	**to**	63

Table 30 Craig: CELF-4 individual subtest scores

2.3.2.2 Speech Measures

A range of formal and targeted wordlists were used to measure speech outcomes. Table 31 gives a summary of the speech measures used in each assessment session. Details of each speech measure are outlined below. All of the speech measures, apart from the questionnaires, were recorded with simultaneous ultrasound. See chapter 4 for the method of ultrasound analysis.

	Pre-Study	Baseline	Pre-VAM	Post-VAM	Pre-UVBF	Post-UVBF	Maintenance	Post-Study
Formal Speech Measures								
GOS.SP.ASS'98	X							X
DEAP Phonology		X	X	X	X	X	X	
Target-specific wordlists								
Untreated /n/		X	X	X	X	X	X	
Treated /n/				X	X	X	X	
Additional Alveolar Wordlist		X	X	X	X	X	X	
Questionnaires								
ICS		X	X	X	X	X	X	
Parent-Questionnaire				X		X		
Child-Questionnaire				X		X	X	

Table 31 Summary of Craig's Speech Measures

2.3.2.2.1 Formal Speech Measures

The Phonology subtest of the DEAP *(*Dodd et al. 2002) was chosen to provide a PCC score and to measure changes in Craig's phonological system over the course of treatment. The DEAP Phonology subtest was recorded at all six assessment time-points (Assessment 1-6) using synchronised ultrasound (see sub-section 2.1.5), audio and lip camera data.

The suitability of the DEAP phonology subtest for assessing velars was carried out by counting the number of velar tokens. Overall, the subtest contains 21 tokens of velars (/k/ = 12, /g/ = 5, /ŋ/ = 4), making up only 15% of the overall number of consonants (141) that are scored to obtain a PCC score. There are also 23 alveolar tokens, making up 16% of the overall number of consonants. Fifteen of these are alveolar plosives and eight are alveolar nasals.

2.3.2.2.2 Target-specific Wordlists – materials and protocol

2.3.2.2.2.1 Untreated wordlist

The untreated velar wordlist consists of 36 single words containing 41 tokens of velar plosives and velar nasals in (mostly singleton) word initial, (mostly intervocalic) medial and (mostly singleton) final positions in a variety of vowel environments. A range of picturable monosyllabic and polysyllabic words were used and each word was elicited through a picture naming task. Photographs from Google images were used for picture naming tasks. Pictures were presented to Craig in blocks of three. If Craig was unable to name the word spontaneously a semantic cue was provided followed by direct imitation if this failed to elicit the desired response. Six sentences were also recorded which included a sample of the 36 single words in connected speech. Sentences were elicited through an imitation task. Table 32 presents the wordlist with sentences, organised into word positions, vowel environments, clusters and sentences. The vowel choices are appropriate for the accent of the child; however, where there are gaps in vowel environments this has been highlighted in grey. See section 2.1.10 for the protocol for scoring the untreated wordlists to obtain a PTCC score.

Level	Vowel	Untreated /k/	Untreated /g/	Untreated /ŋ/
WI Single Word	/i/			
	/e/	cage	gate	
	/ɛ/			
	/a/	car, carrots	gas	
	/ɔ/			
	/o/	comb	goat	
	/u/	cookie		
	/ʌ/	computer, cup	gum	
	/ɪ/		guitar	
	/ə/		gorilla	
WM Single Word	/i/			
	/e/			
	/ɛ/	necklace	lego	
	/a/	jacket	magnet, angry	banging
	/ɔ/			
	/o/			
	/u/	cookie	sugar	
	/ʌ/	bucket	nuggets	
	/ɪ/			singer
WF Single Word	/i/			
	/e/	snowflake		
	/ɛ/			
	/a/	snack	flag, handbag	
	/ɔ/		jog, warthog	strong
	/o/	smoke		
	/u/			
	/ʌ/			
	/ɪ/	magic		jumping, skiing, ring
Clusters		skiing	angry	kangaroo, angry
Sentences		Katie drank her cup of tea. Kris has a kitkat for snack. Chloe came second in the skiing competition.	Gavin played a song on his guitar. Grace liked to build blocks of lego.	Kai is a very good singer.

Table 32 Untreated Velar Wordlist organised into word positions, vowel environments, clusters and sentences

Prior to starting therapy, Craig's PTCC score for single words on the untreated wordlist was 22% at baseline (Assessment 1) and remained relatively stable at 26% in the pre-VAM session (Assessment 2).

2.3.2.2.2.2 Treated wordlist

A treated velar wordlist provided materials for use during intervention (see Table 33). This wordlist consists of words containing velar plosives /k/ and /g/. Treatment materials were modified throughout therapy block one in response to developing client needs. As this wordlist had not been designed in advance, it was not possible to record it in the baseline or pre-VAM sessions.

Craig was able to produce the /ŋg/ sequence in WM position in the untreated wordlist in “kangaroo” and “angry” (see sub-section 2.3.3 for results), therefore therapy began using this as a facilitative environment to elicit /g/. As Craig was able to say “angry”, the /gɹ/ cluster was also included in the treated wordlist, as well as singleton /g/ in WI and WF position at single word and phrase level. /k/ was treated in the later sessions of therapy block one, thus only CV level is included within the treated wordlist. This wordlist was recorded post-VAM only using ultrasound with simultaneous audio and lip-camera data.

Level	Vowel	Treated /k/	Treated /g/	Treated /ŋ/
Isolation		k		
CV	/i/	key		
	/e/	kay		
	/ɛ/	keh		
	/a/	ka		
	/ɔ/	kaw		
	/o/	co		
	/u/	coo		
	/aɪ/	Kai		
	/ɪ/	kih		
WI single word	/i/			
	/e/		game	
	/ɛ/		gecko	
	/a/			
	/ɔ/		golf	
	/o/		ghost, go, goldfish, goal	
	/u/		goose, good	
	/ʌ/			
	/ɪ/		girl, gift	
WF single word	/i/			
	/e/			
	/ɛ/		leg	
	/a/		tag	
	/ɔ/		frog, log	
	/o/			
	/u/			
	/ʌ/		mug, earplug, slug	
	/ɪ/		wig, pig	
WM Clusters			graph, grizzly, green, granny, grow, grumpy, grey, grapes, grin, glue	finger, jungle, tango, longer, stronger, hungry
Phrases			Greg gets the goose Greg gets the gecko Greg gets the frog Greg gets the pig	

Table 33 Treated Velars Wordlist (post-VAM) organised into word positions, vowel environments, clusters and sentences

2.3.2.2.2.3 Treated Alveolar Wordlist

As Craig had reached his velar target in therapy block one with Speech Trainer 3D, therapy block two also introduced alveolar /t/ as an additional therapy target. The treated alveolars wordlist was devised and modified during therapy block two and consisted of CV, VC and VCV syllables, and single words containing mostly singleton /t/ in WI, WM and WF positions (see Table 34). The full wordlist was recorded post-UTI using ultrasound with simultaneous audio and lip-camera data. In the maintenance session, only the highlighted words in Table 34 were elicited and scored due to time constraints within the session and a decrease in Craig's motivation to continue with the recording. Similar to the velar wordlist, it was not possible to record this wordlist pre-therapy as it was modified throughout the therapy block to suit Craig's needs.

Level	Vowel	Treated /t/
Isolation		t
CV	**/i/** **/a/** **/o/** /u/ /aɪ/	tea ta toe two tie
VC	**/i/** **/a/** **/o/**	eat at oat
VCV	**/i/** **/a/** **/o/**	eetee ata oto
WI Single Word	**/i/** **/e/** **/ɛ/** **/ɔ/** **/o/** **/ʌ/** **/ɔɪ/** /aɪ/	teeth tail ten, tent tall toes, toast tongue toys tights
WF single word	**/ɛ/** **/a/** **/ɔ/** **/o/** **/u/** **/ʌ/** **/aɪ/**	pet fat, cat, hat, bat cot, hot boat, float foot, boot nut kite
WM single word	**/i/** **/a/** **/ɔ/** **/u/** **/ʌ/** **/ɪ/** **/aɪ/**	sweetie patting spotty, naughty, potty shooting cutting pretty, hitting, sitting, knitting nighty, writing, biting, fighting
Multisyllabic		t-shirt, teddy, table, tiger, towel, tummy, toddler
Clusters		fort, aunty, dirty, party

Table 34 Craig: Treated /t/ Wordlist (post-UVBF and maintenance) organised into word position and vowel environments. Shaded areas indicate words recorded in maintenance session.

During therapy, there was a focus on WI production as production of WM and WF would typically be produced as a glottal stop in Craig's dialect (Reid 1978; Marshall 2001). This was verified through observation of his twin sister. Therefore, mostly WI tokens were recorded in the maintenance session.

2.3.2.2.3 Intelligibility Measure

The ICS (McLeod et al. 2012) was completed. See sub-section 2.1.7.3 for information on how the ICS was scored.

2.3.2.2.4 Post-Therapy Questionnaires

See method section (2.1), for information on the post-therapy questionnaires. The post-therapy questionnaire for parents was completed in the post-VAM and post-UVBF assessment sessions. The post-therapy questionnaire for children was completed by Craig and the SLT (the author) in the post-VAM and post-UVBF sessions and the 3-month post-therapy questionnaire was completed in the maintenance session.

2.3.2.3 Therapy

2.3.2.3.1 Therapy Block One: VAM

Therapy block one followed the same approach as outlined for Andrew (as in section 2.1.9) and also used Speech Trainer 3D (Smarty Ears 2011). Therapy block one consisted of eight one-hour long sessions, targeting the production of velar plosives, with a focus on the voiced velar plosive /g/ for the most part and bringing in /k/ as a target within the final few sessions. Similar to Andrew's therapy protocol, before targeting the production of velars, the first and second therapy sessions focused on using Speech Trainer 3D to show Craig the different parts of the vocal tract (using labels such as uvula, hard palate, soft palate and "voice box" to explain voicing) and to demonstrate and label sounds of English. During these sessions, auditory discrimination tasks using the iPad and visual discrimination tasks using still frames (screen shots) and dynamic videos from the Speech Trainer 3D model were carried out, for example through games of *same/different* or *pairs.* Craig was asked to describe a sound by identifying if it was a *front* (alveolar) or *back* (velar) sound, a *mouth* (oral) or a *nose* (nasal) sound and a *loud* (voiced) or a *quiet* (voiceless) sound, which he was able to do by the end of session two.

Within his first session, when naming areas of the vocal tract and looking at the back of his mouth in the mirror and identifying that he had no uvula, Craig said that he was unable to make a /ɡ/ sound because he doesn't have a soft palate, indicating good awareness of speech production though some slight terminological confusion (Craig's uvula had been removed during surgery). The treating SLT (tSLT, the author) explained to Craig, using the videos on Speech Trainer 3D to demonstrate, that he does have a soft palate and explained how to make a velar plosive. By his second session, he was able to identify most areas of the vocal tract on the Speech Trainer 3D model, apart from the uvula and the hard palate, which he forgot the names of. Throughout the first block of therapy, a review of the model was carried out at the beginning of every session before production practice, so that Craig could refer back to the app for a visual reference when trying to achieve velar placement. He was able to explain what he had to do with his own vocal tract to produce a velar, e.g. he was able to say that he had to keep the front of his tongue down and keep the back of his tongue up against his soft palate. Therapy followed the same hierarchy as for Andrew.

Table 35 provides information on which level within the hierarchy was worked on during each of the eight therapy sessions. The target singleton or cluster for each level within each session is marked within the table. Both auditory and visual discrimination was also worked on during therapy sessions one and two (indicated with a cross).

Therapy Session	Discrimination	0	1-CV	1-VC	2-WI	2-WF	3	4	5	6	7
1	X	/ŋg/									
2	X	/ŋg/									
3		/ŋg/	/ŋg/	/ŋg/						/gɹ/	
4		/g/					/gɹ/			/gɹ/	
5							/gɹ/	/ŋg/		/gɹ/	/ŋg/
6					/g/	/g/	/g/, /ŋg/			/gɹ/	/ŋg/ /gɹ/
7					/g/	/g/	/g/				
8		/k/	/k/		/g/	/g/					/g/ /ŋg/ /gɹ/

Table 35 Level and target worked on in each therapy session in Craig's therapy block one

As Craig was able to produce [ŋg] within the untreated velars wordlist at baseline and pre-VAM, this was used as a facilitative environment to elicit singular /g/ initially. However, Craig was difficult to engage in tasks targeting singular /g/ and he was unable to elicit /g/ without it having the facilitative environment of a velar nasal, therefore the focus of therapy during session three was to target /ŋg/ in WM position for consistency. As Craig was also able to produce [ŋgɹ] in "angry", the /gɹ/ cluster was also introduced in session three. By session six, Craig was able to produce singleton [g] and progressed onto sentence level for [gɹ] and [ŋg] within the session. By session eight, Craig was able to produce singleton [g], [gɹ] and [ŋg] in connected speech and was able to elicit [k] in isolation and in CV syllables, with a range of vowels.

During each therapy session, Craig's production practice was recorded in order for him to monitor his own progress and to identify his own errors using Speech Trainer 3D as an aid to demonstrate the correct and incorrect productions of his errors. This was particularly useful for Craig, who was reluctant to cooperate with activities if he felt that he would not succeed. By playing back recordings, this allowed Craig to hear his own correct productions and in turn he was more willing to cooperate with more difficult tasks. Frequently during sessions, both the SLT and Craig would score correct productions, using knowledge of results (KR) feedback and would compare results (see Figure 23 for an example). From this example, it is clear that Craig had a good level of awareness of his own errors. Only three tokens were transcribed differently.

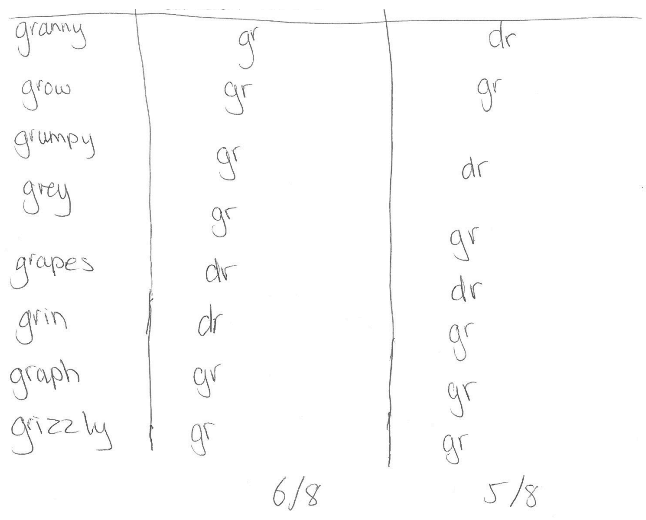


Figure 23 Example of a comparison of the tSLT and Craig's transcription of recorded data throughout a therapy session

2.3.2.3.2 Therapy Block Two: UVBF

Pre-UVBF (assessment 4), Craig scored 76% velars correct (see below for results). The therapy target for therapy block two began as /k/, aiming for /k/ to be more consistent at higher levels on the therapy hierarchy. However, with good success with velars, the ultrasound therapy target changed to /t/ in therapy session three, although generalisation of /k/ continued in the table-top activities without ultrasound. Targeting /t/ also allowed for reinforcement of a /t/-/k/ contrast, which Craig did not have previously, as both were realised as [ʔ]. Similar to therapy block one, the therapy approach used in block two was an articulation approach using the principles of motor learning. Speech Trainer 3D was replaced with UVBF.

The first two therapy sessions in block two focused on learning to associate the movement of the ultrasound video on the screen. As Craig was able to produce velar plosives correctly after therapy block one (although variable), his best velar production was used as a reference for his /k/ target. As he was able to produce [d], this was used as an alveolar target. Parallel to Andrew, ultrasound was used as a visual articulatory model and ultrasound images were compared to those in Speech Trainer 3D through table top activities such as picture matching of static ultrasound and Speech Trainer 3D images and matching of dynamic ultrasound to the dynamic videos in Speech Trainer 3D. A portable ultrasound machine was used by the tSLT to model dynamic ultrasound videos for comparison to the Speech Trainer 3D videos. Ultrasound tasks focused on Craig's ability to use and understand the element of biofeedback, by moving his tongue to target areas on the screen when given gestural instructions from the SLT (e.g. "move the front of your tongue up to this bit of the screen", or "keep the front of your tongue down to make a 'k' sound"). Craig was also instructed to copy static images of his own data (his best production of a velar plosive and /d/) as a reference for a velar tongue shape and an alveolar tongue shape, in order for him to use the biofeedback element of ultrasound to manipulate his tongue to produce a voiceless alveolar and velar plosive. Both tongue shapes were used as a reminder of the contrast alveolar and velar placement for the remainder of the therapy block.

After session two, Craig was able to produce velar plosives using ultrasound and so the target switched to /t/, which he produced with glottal reinforcement. However, activities targeting /k/ in connected speech continued without ultrasound during table-top activities. As discussed above, children in Scotland will frequently replace WM and WF /t/ with a glottal stop. Therefore, therapy focused mostly on WI position, including words with WM and WF tokens in later therapy sessions (6-8) for generalisation into different word positions. By the end of session eight, Craig was able to produce WI, WM and WF /t/ at single word level (mono and polysyllabic words) with 80% accuracy. Table 36 shows the level and target during each session in therapy block two.

Therapy Session	Discrimination	0	1-CV	1-VC	2-WI	2-WF	3	4	5	6	7
1	X	/t/ /k/									
2	X	/t/ /k/									/k/
3	X	/t/	/t/								/k/
4			/t/	/t/	/t/		/t/				
5					/t/		/t/	/t/			/k/ /t/
6				/t/	/t/	/t/	/t/				
7					/t/	/t/	/t/				
8					/t/	/t/	/t/				

Table 36 Level and target worked on in tach therapy session in Craig's therapy block two

2.3.3 Results

2.3.3.1 GOS.SP.ASS'98

The GOS.SP.ASS'98 was repeated by the CLP SLT nine months after Craig's three-month maintenance session, at age 7;6. Results from the GOS.SP.ASS'98 showed that Craig had generalised and maintained the correct place of articulation for velar plosives, velar nasals and alveolar plosives, however there appears to be more evident nasal turbulence on high pressure consonants than in the GOS.SP.ASS from pre-therapy. During this assessment, the CLP SLT noted that Craig presented with mild hypernasality, mild-moderate nasal turbulence and a mild facial grimace. Table 37 summarises the GOS.SP.ASS results pre- and post-study for comparison, with green highlighted areas indicating sounds present in Craig's inventory post-therapy.

	Two Months Pre-Study																		
	Labial					Alveolar						Post-Alveolar			Velar			Glottal	
	m	**p**	**b**	**f**	**v**	**n**	**l**	**t**	**d**	**s**	**z**	**ʃ**	**ʧ**	**ʤ**	**ŋ**	**k**	**g**	**h**	**θð**
SIWI		ʔ͡p	b͌	(f)͡ʔ	ɔ			ʔ	n	n͡ʔ	n	n͡ʔ	ʔ	n		ʔ	n		ð
SFWF		p͡ʔ	b	f͌	v			ʔ	n	s↓		s̰↓	(t)͡ʔ	ħ	n	ʔ	n		
	Nine Months Post-Study																		
	Labial					Alveolar						Post-Alveolar			Velar			Glottal	
	m	**p**	**b**	**f**	**v**	**n**	**l**	**t**	**d**	**s**	**z**	**ʃ**	**ʧ**	**ʤ**	**ŋ**	**k**	**g**	**h**	**θð**
SIWI						n̠		t͌		s͌	z	s͌ ʲ	ts	dz			ɡ͌		ð
SFWF						n		t͌		s	z͌	ʃ	ts	dz			g		

Table 37 Craig's GOS.SP.ASS'98 Consonant Production Pre-and Post-Study. Green shading indicates sounds present in Craig's speech inventory.

2.3.3.2 DEAP Phonology

Figure 24 shows Craig's PCC scores from the DEAP Phonology subtest. Craig's PCC score was stable across baseline and Pre-VAM at 52%, with a slight increase post-VAM at 59%. This decreased marginally pre-UVBF to 56%. Post-UVBF Craig's PCC score had increased by 21 percentage points to 76% with a further increase in the maintenance session at 84%.

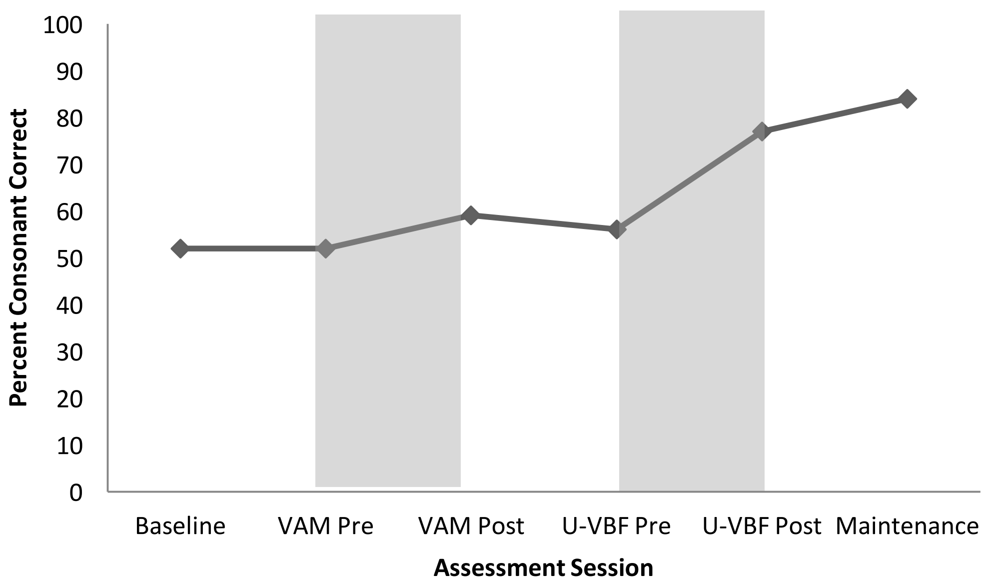


Figure 24 Craig's DEAP Phonology subtest PCC Scores. Grey shading indicates periods of intervention

Table 38 provides details of the errors identified in the DEAP data, including typical patterns and atypical error patterns. These atypical errors, grouped into the following cleft-type characteristics: retraction to glottal, lingual retraction, lingua-glottal double articulations, supralaryngeal double articulations, and idiosyncratic errors. These make up between around 30% and around 50% of Craig's errors in the DEAP across each of the six sessions.

NON-CLEFT PROCESS	**Baseline**	**Pre-VAM**	**Post-VAM**	**Pre-UVBF**	**Post-UVBF**	**Maintenance**
Velar Fronting	4	3	2	1		
Post alveolar fronting	3	4	5	9	8	5
Dentalisation of alveolar fricatives	7	4	5	5	2	
labialisation of fricatives	1			2		
Backing*	11	13	5	10	3	
Gliding				2	1	
Stopping	4	4	3	1		
Deaffrication	1	2	1	1		1
Voicing errors	9	2	12	7	3	3
FCD		1				
ICD	1	5		1	4	2
MCD	1	1		2		2
Cluster Reduction	4	8	7	5	3	2
Other*	21	20	18	16	10	7
TOTAL	67	67	58	62	34	22

***BACKING/OTHER: SPECIFIC ERRORS**	**Baseline**	**Pre-VAM**	**Post-VAM**	**Pre-UVBF**	**Post-UVBF**	**Maintenance**
/g/ - [ʔ]	2	2				
/k/ - [ʔ]	5	4		3		
/t/ - [ʔ]	3	2	5	5	3	
/s/ - [ʔ]	4	1	1	1		1
/θ/ - [ʔ]	1		1	1	1	
/k/ - [k͡ʔ]	2	2	1		1	
/k/ - [d͡ʔ]		1	2			1
/p/ - [p͡ʔ]	2	2		1		
/t/ - [d͡ʔ]	1	1	1	1		
/t/ - [t͡ʔ]		1		1	1	
/ð/ - [l]	1		1	1	1	
/t/ - [n]	2			2	1	
/k/- [n]				3		
silent articulation	1		2	3	2	
Other (idiosyncratic)	8	17	9	4	3	5
TOTAL Backing and Other	32	33	23	26	13	7
% Total Errors	48%	49%	40%	42%	38%	32%

Table 38 Craig's DEAP Error Pattern Analysis, separated into non-cleft processes and specific errors associated with CP, such as retraction and double articualtions, and idiosyncratic errors

2.3.3.3 Treated Wordlists

Post-therapy block one (Post-VAM), Craig scored 95% velars correct in the treated velars wordlist. Post-therapy block two (Post-UVBF), he scored 54% /t/ correct in the treated alveolars wordlist. In the maintenance session, he scored 76% /t/ correct on the treated alveolars wordlist; however not all of this wordlist was elicited and only those highlighted in Table 34 were scored. There was however no baseline for either the treated velar or alveolar wordlist, therefore improvement is based on the DEAP phonology subtest and the untreated velar wordlist.

2.3.3.4 Untreated Velar Wordlist

Figure 25 shows Craig's PTCC scores obtained from the 36 single words (41 velar tokens) and six sentences (24 tokens) across all six assessment time-points, transcribed by the treating SLT (grey highlighted sections indicates periods of therapy). When looking at the single words PTCC scores, at baseline Craig achieved a PTCC score of 22%, which remained relatively stable, although slightly higher, in the pre-VAM assessment with a score of 26%. In the Post-VAM assessment Craig's PTCC had increased to 76%, with correct productions of [ŋ], and velar plosives in all word positions, lower than the 95% correct in the treated wordlist. Scores remained stable over the inter-therapy break. After the second block of therapy this had risen to 93% in the Post-UVBF session, which remained relatively stable, although slightly lower, at 90% in the maintenance session. Preston et al. (2014) suggests that 20% improvement is clinically significant, therefore Craig's PTCC results show a clinically significant improvement post-VAM at single-word level.

Twenty-four tokens of velars were scored in the sentences. At baseline, Craig had a PTCC score of 21%, which decreased to 13% pre-VAM. Post-VAM the PTCC score increased to 50%, which remained stable through to Pre-UVBF. However, it should be noted that pre-VAM, only three sentences were recorded due to Craig's cooperation with the recordings. Post-UVBF the PTCC score in sentences had increased to 83%, with a slight increase to 88% in the maintenance session. An increase from 21% to 88% overall also indicates a clinically significant increase in connected speech (Preston et al. 2014).

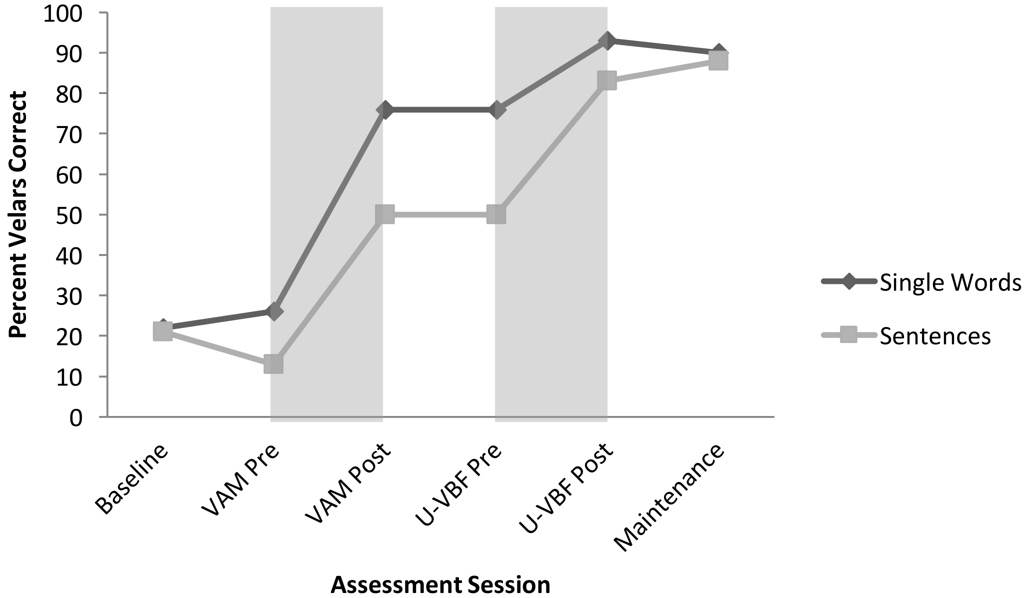


Figure 25 Craig's PTCC Scores from the Untreated velar wordlist (single words and sentences). Grey shading indicates periods of intervention

2.3.3.4.1 Phonological Environment Analysis

Incorrect productions of velars in the untreated wordlist were scored depending on the segmental errors, e.g. if /k/ was produced as [d͌] an error was scored for velar fronting, voicing error and accompanying VP friction, i.e. the number of errors does not correspond with the number of velar tokens in the wordlist. Most errors were found in word initial (N=63) with fewest errors found in WF position (N=27). The highest number of correct productions was found in word medial position (N=59) with 23 of these correct productions being velar nasals and 35 being velar plosives. The highest number of correct velar plosives was found word initially (N=47). Velar fronting was most prevalent in WI position, which decreased over time. In the Post-VAM session, velar fronting was only found in WF position. It is clear from the three figures below that the number of correct velar plosives and nasals increases over the course of the six sessions, with errors on velars being variable regardless of word position.

Figure 26 shows the number of correct velar plosives across the six assessment sessions in WI position. At baseline and pre-VAM, Craig had no correct WI velar plosives, with an increase to 12 Post-VAM. The most correct tokens (N=13) found in the post-UVBF session.

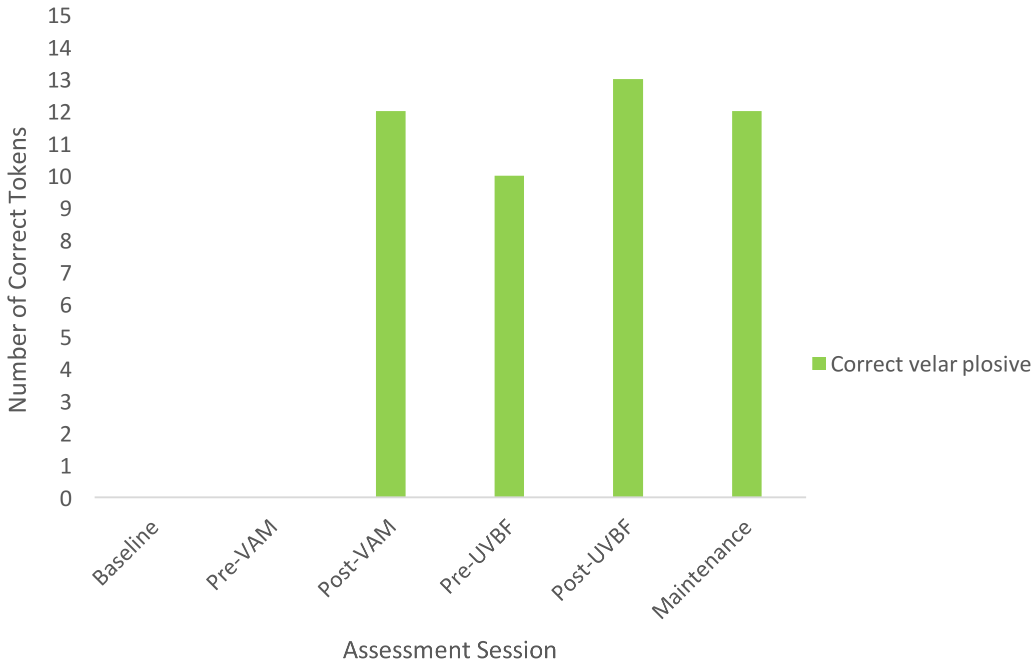


Figure 26 Craig: Untreated Wordlist Correct Tokens (WI Position)

In baseline and pre-VAM, there were two correct WM tokens of velar plosives (Figure 27). Similarly this increased post-VAM to seven correct tokens, with the most correct WM tokens found in maintenance (N=9). The number of correct velar nasals remained stable throughout, with one incorrect production in pre-UVBF.

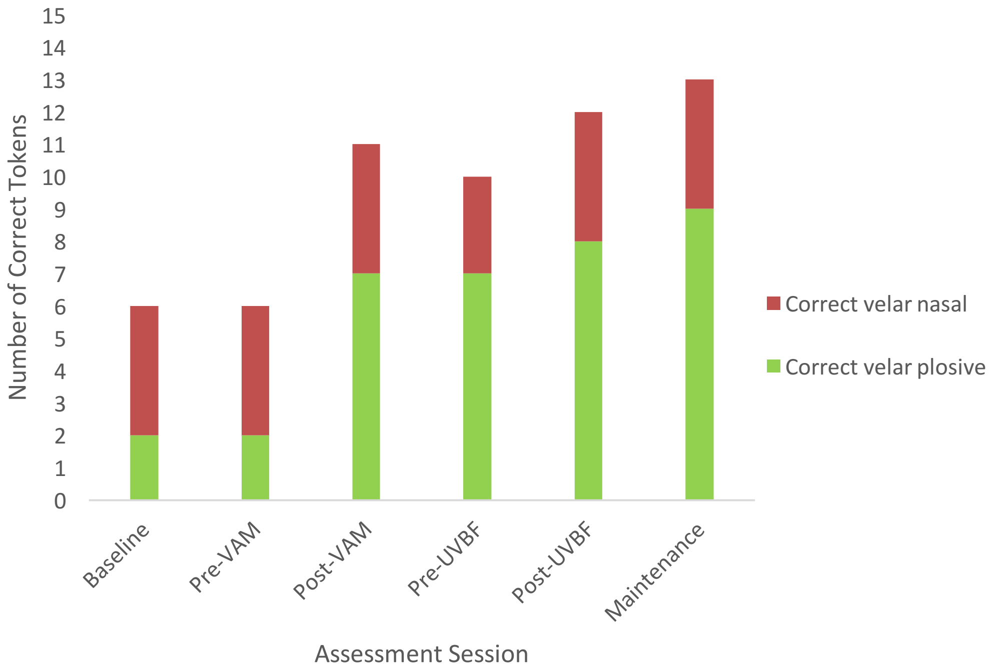


Figure 27 Craig: Untreated Wordlist Correct Tokens (WM Position)

Similar to WI position, there were no correct tokens of velar plosives in WF position (see Figure 28). This increased to five correct tokens post-VAM, which further increased to eight post-UVBF and remained stable through maintenance. Velar nasals remained relatively the same throughout all six sessions.

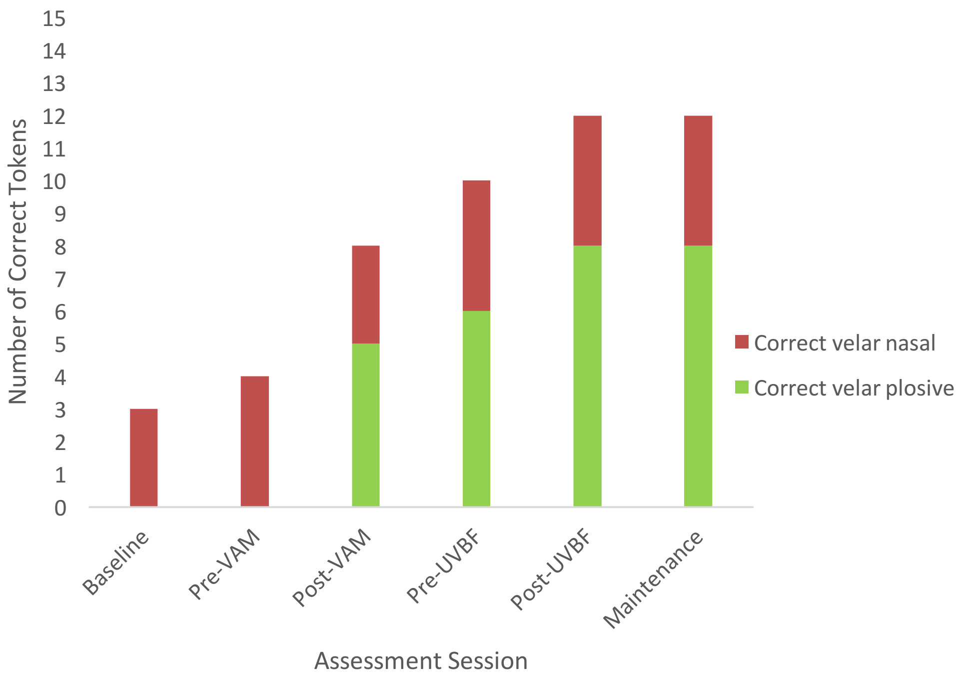


Figure 28 Craig: Untreated Wordlist Correct Tokens (WF Position)

Figure 29 shows the error pattern analysis for WI position. It is evident that errors were very variable, particularly in baseline and pre-VAM sessions, and that there were multiple errors on single tokens (see example above). Out of a possible 41 tokens of velars, velar fronting was the most common error in baseline (N=8) and pre-VAM (N=7), which reduced to no velar fronting in post-VAM. There were two further sessions with velar fronting, however substantially less so than in baseline and pre-VAM. Retraction to glottal was also found in baseline (N=3) and pre-VAM (N=5), with voicing errors also prevalent pre-therapy. Retraction to glottal and voicing errors had eliminated post-UVBF. It is evident from the three figures below that word position had an effect on the number of errors, with WI having more errorful tokens in the baseline and pre-VAM sessions than WM and WF position. WI position also had the most accompanying VP friction, with five out of six sessions having at least one token with VP friction.

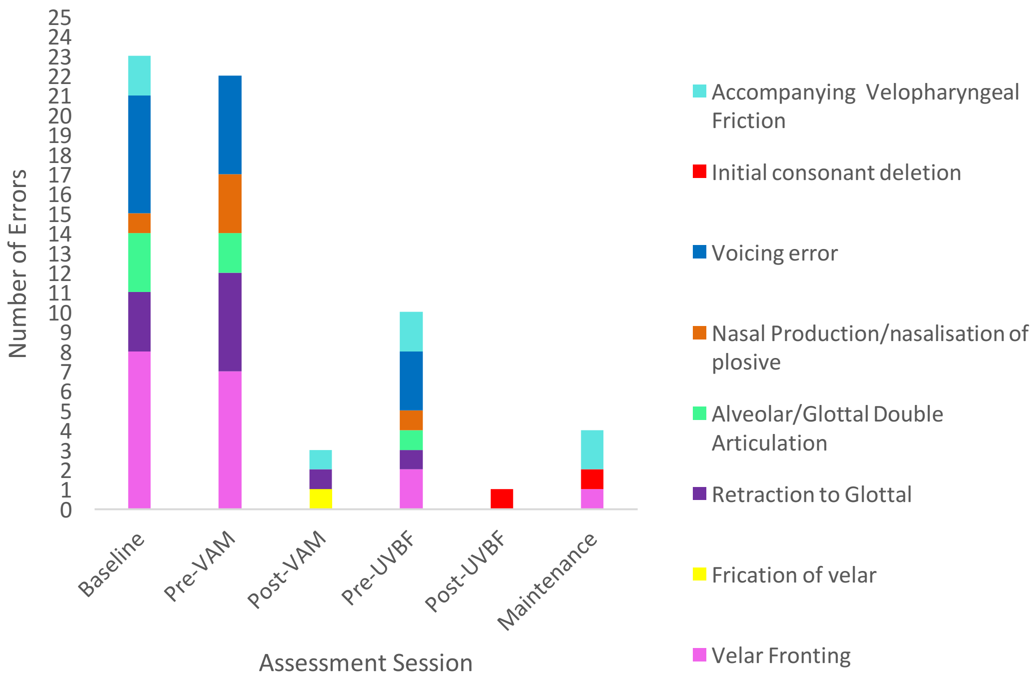


Figure 29 Craig: Untreated Wordlist Error Pattern Analysis (WI position)

Velar fronting was also evident in WM position (see Figure 30), which was eliminated post-VAM. Retraction to glottal remained throughout the sessions, although decreased from baseline (N=1) to maintenance (N=1). Similar to WI position, there are high levels of variability in Craig's production of velars in WM position. There were only two tokens with accompanying VP friction in baseline and maintenance.

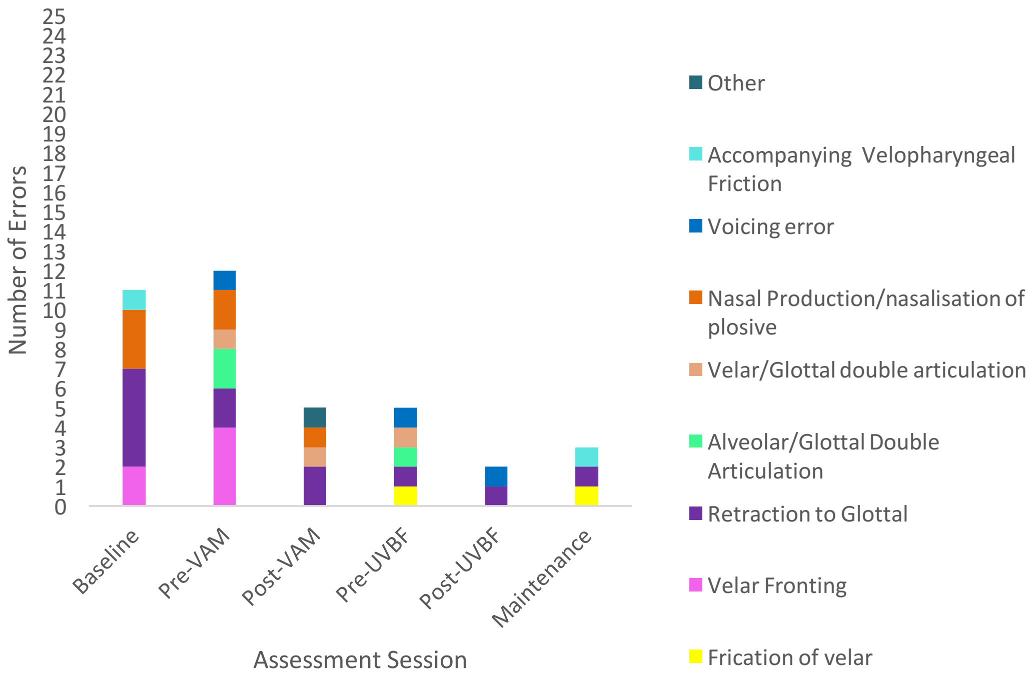


Figure 30 Craig: Untreated Wordlist Error Pattern Analysis (WM position)

Unlike in WI and WF position, there are fewer errors in WF position, with no WF errors in the pre-UVBF and maintenance sessions. In WF position (see Figure 31), retraction to glottal placement was the most common error, with velar fronting evidence in two sessions (baseline, N=3) and post-VAM, N=1). In three of the sessions, velar plosives had accompanying VP friction.

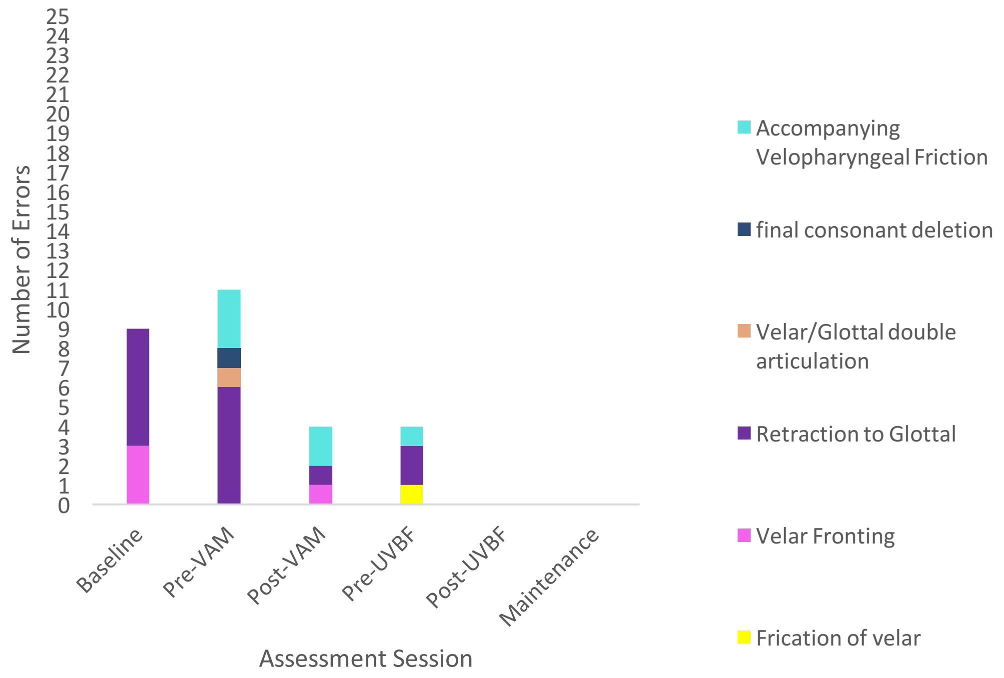


Figure 31 Craig: Untreated Wordlist Error Pattern Analysis (WF position)

2.3.3.4.2 Intra-rater reliability of Untreated velars

The researcher transcribed all 36 single words twice, with a three-year interval between transcription one and transcription two. Results showed agreement across the two time-points (mean=96%, range=93%-98%). Figure 32 shows percentage of agreement across all six assessment time-points.

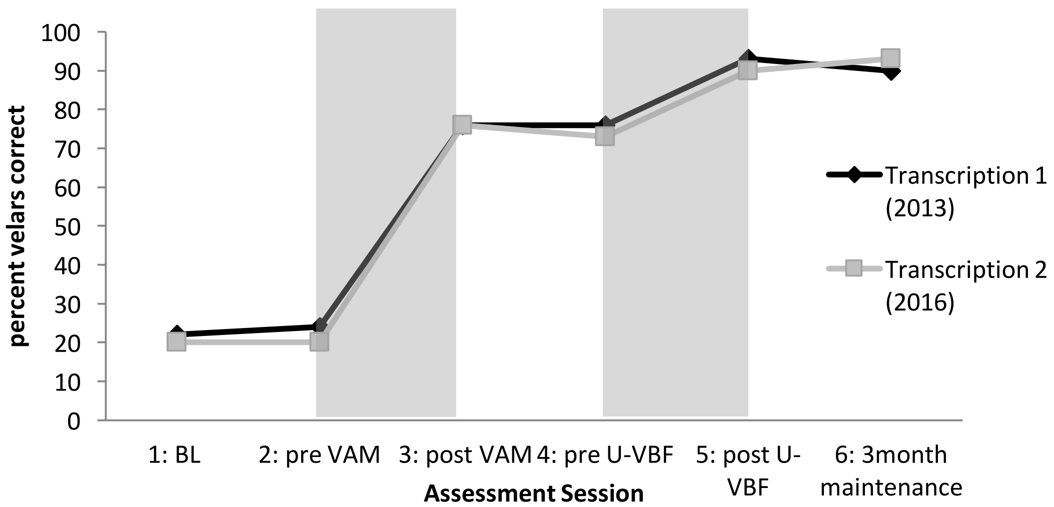


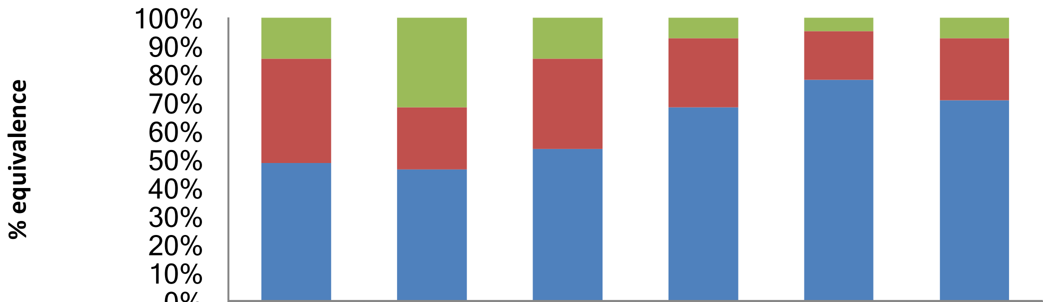
Figure 32 Craig: Intra-Rater reliability percent velars correct scores. Grey areas indicate periods of intervention

Statistical analysis showed that five out of the six sessions had "very good agreement", with one sessions (post-UVBF) having "moderate agreement". Table 39 shows Cohen's Kappa results for all six assessment sessions.

Session	no. observed agreements	no. expected by chance	Kappa Score	SE of Kappa*	95% CI**	Strength
Baseline	40 (97.56%)	27.5 (67.10%)	0.926	0.073	0.783-1.000	very good
Pre-VAM	39 (95.12%)	26.2 (65.62%)	0.858	0.097	0.668-1.000	very good
Post-VAM	39 (95.12%)	24.9 (60.74%)	0.876	0.086	0.708-1.000	very good
Pre-UVBF	40 (97.56%)	25.4 (61.87%)	0.936	0.063	0.812-1.000	very good
Post-UVBF	38 (92.68%)	34.6 (84.36%)	0.532	0.237	0.068-0.977	moderate
Maintenance	40 (97.56%)	34.6 (84.35%)	0.844	0.152	0.546-1.000	very good

Table 39 Craig: Intra-Rater Reliability Cohen's Kappa Scores *SE = standard error; ** CI = confidence interval

Equivalence scores for transcriptions over the two time-points (2013 and 2016) (see sub-section 2.1.10.1) were calculated using a three-point scale (2=same; 1=almost equivalent; 0=different). Figure 33 shows equivalence scores for each assessment time-point and percentages of each equivalence score. Within each session, the majority of transcriptions were the same, with the lowest percentage of tokens being "different", with the exception of pre-VAM, where there was a higher percentage of "different" than "almost equivalent".



	1: BL	2: pre VAM	3: post VAM	4: pre U-VBF	5: post U-VBF	6: 3month maintenance
0: different	6	13	6	3	2	3
1: almost equivalent	15	9	13	10	7	9
2: same	20	19	22	28	32	29

Figure 33 Craig Intra-Rater Equivalence Scores

2.3.3.4.3 Inter-rater reliability of untreated velars

Inter-rater reliability shows that all three transcribers agreed on individual token pairs over 70% of the time across all six assessment sessions (mean = 76% range = 71%-85%, "intermediate to good agreement"). Statistical analysis showed that the highest agreement across transcribers was found in the Pre-UVBF session (Fleiss' Kappa = .7375) with all three transcribers agreeing on 33/41 tokens. The lowest agreement was found in the maintenance session (Fleiss' Kappa = .5597) with all three transcribers agreeing on 29/41 tokens (Roxburgh et al. 2016, see section 7.10). Figure 34 shows PTCC scores derived from all three transcribers.

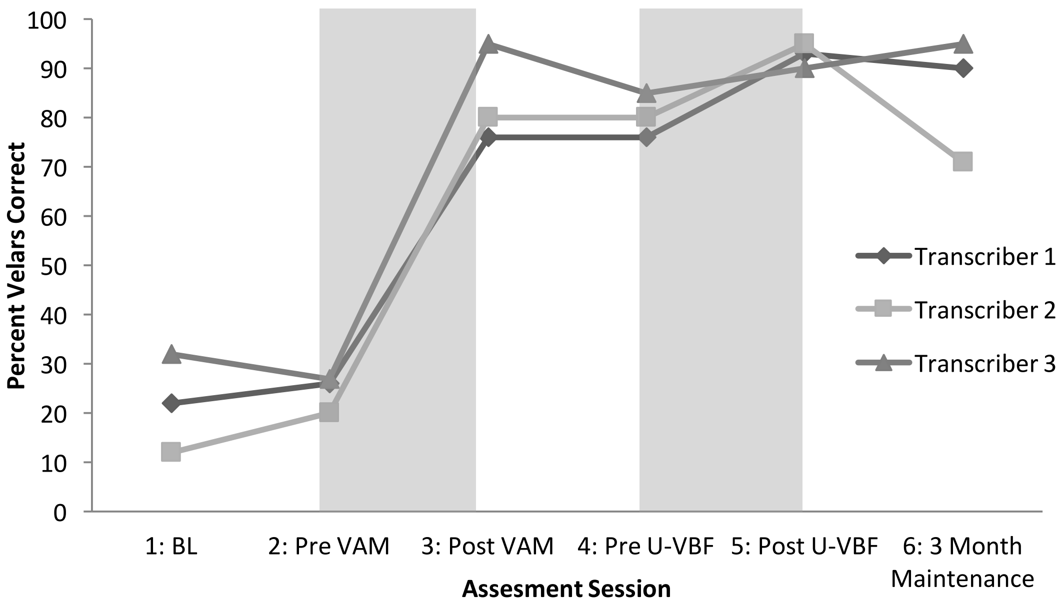


Figure 34 Craig's PTCC Scores obtained from all three transcribers for inter-rater reliability

Whilst agreement was found between the three transcribers, it is important to point out that in the PTCC scores were not always in a similar trend across the three transcribers. For example, between baseline and pre-VAM, transcribers two and three report an upward trend in PTCC scores, while transcriber one (tSLT) reports a downward trend. Similarly, there are two sessions (pre-UVBF and maintenance) where two transcribers report stability and one (tSLT in pre-UVBF and transcriber two in maintenance) reports a decrease in PTCC scores. This highlights the complexity of Craig's speech, which led to reliability scores being lower than the levels expected, based on Shriberg and Lof (1991).

2.3.3.5 Feedback Questionnaires

2.3.3.5.1 Intelligibility in Context Scale

The Intelligibility in Context Scale shows improvement across the five sessions in which it was completed by Craig's Mother (see Table 40). From baseline to post-VAM, there was a change from "sometimes understood" to "usually understood" for immediate family members, with a change in acquaintances from "rarely understood" to "sometimes understood". This remained the same between post-VAM and Pre-UVBF, as there was no therapy between these sessions. Post-UVBF, there was a change from "sometimes understood" to "usually understood" for extended family, friends and acquaintances, with a change in strangers from "rarely understood" to "sometimes understood". In the maintenance session, three categories (parents, immediate family and teachers) changed from "usually understood" to "always understood", with a change in strangers from "sometimes understood" to "usually understood".

	Baseline	Post-VAM	Pre-UVBF	Post-UVBF	Maintenance
Do you understand your child?	4	4	4	4	5
Do immediate members of your family understand your child?	3	4	4	4	5
Do extended members of your family understand your child?	3	3	3	4	4
Do your child's friends understand your child?	3	3	3	4	4
Do other acquaintances understand your child?	2	3	3	4	4
Do your child's teachers understand your child?	4	4	4	4	5
Do strangers understand your child?	2	2	2	3	4
TOTAL	21	23	23	27	31
AVARAGE TOTAL	3	3.3	3.3	3.9	4.4

*** 5 = Always; 4 = Usually; 3 = Sometimes; 2= Rarely; 1 = Never**

NB. ICS was not completed Pre-VAM as it was only 1 week after Baseline

Table 40 Intelligibility in Context Scale Scores

2.3.3.5.2 Therapy Outcome Questionnaire for Parents: Parental Responses

Post-VAM Therapy: Overall, Craig's Mother reported that his speech had "*greatly improved*". She reported that after therapy block one, his awareness of speech sounds had "*moderately improved",* and his ability to articulate the target sounds "greatly improved". It was reported that post-VAM, he was able to use /k ɡ ŋ/ "*always*" with familiar words, but only "*most of the time*" with unfamiliar words in conversation. Compared to his siblings, his Mother rated his speech as "*much worse*". Her view of Speech Trainer 3D was that it had helped him achieve his therapy goals greatly.

Post-UVBF Therapy: Craig's Mother reported that his speech had "*greatly improved*". His Mother reported that after therapy block two, his awareness of speech sounds and his ability to articulate the target sounds "*greatly improved*". It was reported that post-UVBF, he was able to use /k ɡ ŋ t d/ "*most of the time*" in conversation. Compared to his siblings, his Mother rated his speech as "*slightly worse*". She reported that ultrasound greatly improved Craig's speech as he was able to see his tongue and change his tongue position to create new sounds. She felt that using the ultrasound helped him achieve his therapy goals.

2.3.3.5.3 Therapy Outcome Questionnaire for Children: Participant Responses

Post-VAM Therapy: Craig reported that he felt using the iPad was "good" because he was able to look at the pictures and the words on the screen. He enjoyed using the iPad because he got to use it independently by swiping through the app. He reported that the worst/hardest bit about using the iPad was producing /ɡ/ on its own. He felt that using the iPad has helped his speaking. Table 41 shows Craig's responses to questions regarding intelligibility with a range of listeners. When asked if the sessions were "too short, just right or too long" he reported that they were "just right" (around one-hour).

Post-UVBF Therapy: Craig reported that he thought looking at his tongue was "awesome" because he liked it when his tongue was going backward and forward. He reported that he enjoyed the ultrasound sessions and seeing his own tongue moving. The worst bit about the ultrasound was getting the headset on and the hardest bit was reading all of the prompts. He felt that using the ultrasound helped him learn new sounds. When asked if the sessions were “too short, just right or too long” he reported that they were “too long”. These sessions were one hour, with 30-minutes spent using ultrasound.

	Post-VAM	**Post-UVBF**
How often do you think your parents understand you when you speak?	Rarely	*Almost Always*
How often do you think your brothers or sisters understand you when you speak?	Always	Sometimes
How often do you think your teacher at school understands you when you speak?	Always	Always
How often do you think your friends understand you when you speak?	Never	Never
When you talk to new people, how often do they understand you when you speak?	Never	Never

Table 41 Craig's Responses to Intelligibility Questions

Maintenance: Craig reported that he preferred ultrasound to Speech Trainer 3D. When asked why, Craig said he liked it when his tongue moved on the screen and that it was easier to read than the VAM in Speech Trainer 3D. Craig required additional prompting during this questionnaire to answer questions about why he preferred ultrasound.

2.3.4 Clinical Discussion

Parallel to the clinical discussion for Andrew, the following section will reflect on the use of Speech Trainer 3D and ultrasound visual biofeedback, with reference to the objective PTCC scores derived from transcriptions, parental and child responses to questionnaires and will consider the researcher’s perspective throughout a clinical discussion.

2.3.4.1 Therapy Outcomes

Craig was referred to the current study by the CLP specialist SLT only four-months post-surgery (secondary surgery - Hynes Pharyngoplasty). The CLP SLT completed the GOS.SP.ASS with Craig two months prior to the study commencing. This showed that Craig had a limited phonetic inventory, with only four established consonants [m n l h]. The CLP requested that velar plosives be the therapy target for Craig. From his GOS.SP.ASS, Craig was velar fronting /ɡ/ and /ŋ/ to [n] and retracting /k/ to glottal placement. Alveolar plosives were also incorrect, with /d/ realised as [n] and /t/ retracted to glottal placement. Therefore, therapy targets were velar plosives and alveolar /t/, reinforcing an alveolar/velar contrast. Craig received two blocks of therapy, one using Speech Trainer 3D as a VAM and one using UVBF. Therapy outcomes were based on PCC scores derived from the DEAP Phonology subtest and PTCC scores from untreated wordlists, measured across six assessment sessions.

It was hypothesised that Craig's PTCC scores would remain stable during therapy block one and would increase post-UVBF. PTCC scores at baseline (22%) and maintenance (26%) were stable. Post-VAM, there was a considerable increase in PTCC to 76% with correct productions of [ŋ], and velar plosives in all word positions, contrary to expectations that this would remain stable during therapy block one. Whilst Craig continued to have some input targeting velars during therapy block two, this was mostly during the table-top activities and not with ultrasound biofeedback, therefore the further increase in PTCC post-UVBF, through to maintenance is likely due to generalisation rather than ultrasound.

Whilst there was not an untreated wordlist for /t/ to assess for generalisation across a range of words, the DEAP can be used to measure change in Craig's production of /t/. Pre-UVBF, Craig made various different errors (10) with /t/, including retraction to [ʔ], retracting to /k/, double articulations, or nasalisation. Post-UVBF, there were fewer errors (5), with even fewer (1) in maintenance. Thus, indicating an improvement after using UVBF, although there were only a small number of tokens of /t/ in the DEAP.

Post-study, the GOS.SP.ASS also indicated improvement for both alveolars and velars. Velars /ɡ/ and /ŋ/ were fronted to alveolar placement and realised as a nasal stop [n]. Similarly, alveolar /d/ was realised as a nasal [n]. Velar /k/ and alveolar /t/ were both transcribed as [ʔ]. Whilst there were errors with both alveolars and velars, there was no contrast between the two places of articulation. Both of the nasal stops were realised as alveolar nasal [n], both of the voiceless stops were retracted to glottal placement and both of the voiced stops were produced as [n]. nine months post-study, the GOS.SP.ASS indicated correct place of articulation for all alveolars and velars, with accompanying VP turbulence on /t/ and /ɡ/. Post-therapy, there was a clear contrast between alveolar and velar consonants, indicating generalisation of both of the therapy targets.

Intelligibility scores from the ICS also increased from baseline to maintenance, with parents reporting improvement in intelligibility to all listeners. There was improvement in intelligibility with immediate family members after the first block of therapy, with further improvements post UVBF found in intelligibility with extended family, friends and acquaintances. Within the three-month maintenance period, improvements were reported in intelligibility for parents, immediate family members, teachers and strangers. In the post-therapy questionnaire, Craig also reported that his parents almost always understand him post UVBF; however, he reported that his siblings always understand him post-VAM but only sometimes post-UVBF. He reported that his teachers always understand him post-VAM and UVBF and that his friends and strangers never understand him. This does not correspond with his parent's views in the ICS, however does give an indication of Craig's level of awareness of his own speech difficulties and how listeners perceive his speech. Craig's perception that his friends and strangers are not able to understand him may have impacted on his performance during therapy.

Craig was often difficult to engage in tasks due to poor self-confidence and reluctance to try during activities. Whilst he reported in the three-month questionnaire that he preferred ultrasound to the iPad, Craig would often become upset during sessions in therapy block two. This was mostly due to the headset, which he reported in the post-UBF questionnaire as the worst bit about ultrasound.

2.3.4.2 Difficulties with Phonetic Transcription

Whilst Andrew's data was difficult to transcribe, with very variable productions and a large number of both typical and atypical speech errors, Craig's data was even more challenging. Though inexperienced at the time of the first transcriptions, statistical analysis of suggested "very good" intra-rater reliability for all apart from one session (post-UVBF), with a three-year gap and more post-graduate experience with complex SSDs.

Inter-rater reliability showed over 70% agreement between the tSLT and two experienced phoneticians, which is above the average of 40% agreement found in Gooch et al. (2001) for speech in individuals with CP. However, based on Shriberg and Lof (1991) and Shriberg et al. (1997)'s range in reliability (80% for narrow transcriptions and 90-90% for broad transcriptions), this would suggest poor inter-rater reliability. This supports the literature (e.g. Sell 2005) that despite phonetic transcription being gold standard for speech outcomes for CP, complex errors such as those found in Craig's data, leads to poor reliability and thus, more than one experienced listener should be involved in the assessment of speech in individuals with CP.

2.3.4.3 Evaluation of Speech Materials

Similar to Andrew, there are also obvious flaws in the speech materials for Craig. Untreated wordlists should have included a wider range of vowels and words with varying complexity. One of the major issues with speech materials for Craig was the lack of an untreated alveolar wordlist. As it was not anticipated that Craig would make improvement in therapy block one, only a velar wordlist was devised. However, after the post-VAM recording, an untreated alveolar wordlist should have been included for the pre-UVBF recording rather than the untreated velars wordlist.

Likewise, it would have improved the protocol to have treated wordlists for both alveolars and velars recorded pre- and post-therapy. However, it was unforeseen at baseline that alveolars would be included as a treatment target. Ideally, the additional velar wordlist, which included minimal pairs, would have also been recorded as the additional alveolar wordlist was for Andrew. However, due to time constraints within recording sessions and lack of motivation from Craig to continue with recordings, this was not possible.

2.3.4.4 Evaluation of Therapy Tools

Whilst PTCC scores and questionnaires support improvement in Craig's production of velars, it is difficult to conclude for Craig whether one tool (VAMs or UVBF) had an advantage over the other in his treatment. Craig had clearly acquired his therapy target (velars) after therapy block one, before commencing UVBF therapy. Gibbon and Wood (2010) suggest that VBF is most useful for establishing motor programmes for new articulations, as it provides immediate and delayed KP feedback which is beneficial for acquisition of a new motor skill during pre-practice, therefore UVBF was perhaps unnecessary once Craig had learned to produce a velar using Speech trainer 3D. Whilst /t/ was also targeted using UVBF, there is limited evidence to support this as there was no untreated wordlist to measure pre- and post-therapy outcomes. The DEAP and GOS.SP.ASS however, do show improvement in production of alveolars, with the GOS.SP.ASS indicating generalisation of alveolar placement nine months post-therapy, albeit with some accompanying VP turbulence.

Craig was able to effectively use the iPad app independently. He used it as a VAM for demonstration of place of articulation for velar plosives. By session two, he was able to describe how a velar plosive was made and was able to label the areas of the vocal tract. When listening back to audio recordings of his own productions, Craig was able to scroll through the app to identify his production, whether it was correct or incorrect. If it was incorrect, he was able to describe what he did wrong, using the VAM to point to which part of the tongue he thought he used (e.g. the front of the tongue against the alveolar ridge) and then identify which part of the tongue he should have used and where the point of contact should have been (e.g. the back of the tongue against the soft palate).

When using ultrasound, Craig's image quality made it difficult to interpret live images (see chapter 4). As Craig had a very small chin, tongue tip information was often missing from the data. As his therapy target changed to /t/, this made therapy challenging as Craig was unable to see the tip of his tongue for biofeedback and the tSLT was unable to give accurate instruction and knowledge of performance (KP) feedback during sessions. Craig also did not like wearing the headset, which meant that sessions using ultrasound were often shorter than with Andrew as Craig could not tolerate wearing the headset as long as Andrew could. Craig could be difficult to engage in tasks, therefore potentially leading to poorer therapy outcomes in therapy block two with ultrasound. Despite this, Craig stated in his post-therapy questionnaire that he preferred ultrasound to Speech Trainer 3D as he could see his own tongue moving in real-time.

2.4 Summary of Treatment Study

The current study used a single-subject design and compared two treatments using an ABACA design. Two participants with repaired SMCP, Andrew and Craig, received six assessment sessions and two blocks of therapy, each consisting of eight one-hour sessions, across a nine-month period. Therapy for both Andrew and Craig followed a motor-based therapy approach, using Speech Trainer 3D as a VAM in therapy block one and UVBF in therapy block two. Therapy outcomes were measured by obtaining a PTCC score from target-specific wordlists and PCC score from the DEAP Phonology subtest. Results from the GOS.SP.ASS'98, completed by the CLP specialist SLT two to four months before enrolling on, and around nine-months after the therapy ended, were also used to measure generalisation and maintenance of their therapy targets (Andrew, /n/; Craig, velars). Intelligibility was measured using the ICS and feedback questionnaires were completed by parents and children after each block of therapy, with a further questionnaire completed by children at their three-month maintenance assessment.

For both children, PTCC scores derived from phonetic transcription improved from baseline to maintenance (three months after therapy ceased). For Andrew, this improvement was modest, rising from 5% PTCC at baseline to only 21% PTCC at maintenance. However, there was an upward trend in PTCC scores overall. In line with Preston, et al. (2014) this would suggest that scores do not represent a clinically significant improvement in his production of /n/, thus implying that neither therapy was particularly effective for Andrew. In contrast, Craig improved from 22% PTCC at baseline to 90% PTCC at his maintenance recording, suggesting he had successfully integrated velars into untreated words. However, velars were mostly targeted with VAMs and not with UVBF, therefore the agents of change cannot be deciphered. In addition, /t/ was also targeted for Craig, and there were obvious gaps in the speech materials, with no untreated wordlist to compare pre- and post-therapy. However, results from the DEAP and from the GOS.SP.ASS, completed by the CLP specialist SLT, indicate that Craig had generalised (learned) the new motor plan for both alveolar and velar consonants, which Craig maintained nine months after therapy on the project concluded.

Though the ICS did not indicate any changes in intelligibility for Andrew, it showed an increase in intelligibility to all listener types for Craig. In contrast, Craig reported post-therapy that his friends and strangers did not always understand him, possibly due to low levels of confidence. Despite both children achieving more progress in therapy block one; both of them reported in the three-month questionnaire that they preferred ultrasound as they were able to see their own tongue moving. Whilst the current study did not investigate each of the children's ability to innately tongue-read, it is interesting that Craig also reported that ultrasound was easier to "read" than the VAM. Previous studies have investigated the ability to tongue read talking heads (Badin et al. 2010) and compared ultrasound with EPG (Cleland et al. 2013); however, none have compared tongue reading abilities for ultrasound and VAM. Future studies should investigate and compare children's ability to tongue read in both VAMs and ultrasound, which would in turn lead to a better understanding of whether ultrasound has any advantage over VAMs and why.

There were challenges in phonetic transcriptions for both Andrew and Craig, in particular. Previous literature suggests that point-by-point reliability for broad phonetic transcription is often in the 90-95% range and for narrow transcription is often around 80% (Shriberg and Lof 1991; Shriberg et al. 1997). However, Preston et al. (2011) note that complex speech disorders, such as those found in CP, are often associated with low inter-rater agreement due to their complexity. When comparing listener judgements against transcriptions of compensatory articulations in speakers with CP, Gooch et al. (2001) found an average of 40% agreement across listeners (range 19%-71%). Based on the range of reliability proposed by Shriberg and Lof (1991) and Shriberg et al. (1997), results would suggest that our average of 74% accuracy across both speakers is not reliable, thus suggesting the need for further investigation into the assessment data, such as conducting a multi-listener perceptual evaluation and articulatory analysis. Whilst it is recognised that perceptual assessment should include multiple listeners, Howard (2004) also supports the need for instrumental analysis.

When evaluating the therapy protocol, it was felt that the protocol was perhaps not as rigid as it should have been. A protocol such as that in Preston et al. (2014) and Cleland et al. (2015c) would be used for any future studies. If this were in place for the current study, then both Craig and Andrew's productions would have been more consistent at each level before moving on to the next level in the articulation hierarchy when the previous level was not yet learned or generalised. This may have affected PTCC scores overall. Therapy targets for Craig would also have been reconsidered should this study be conducted again. Untreated wordlists would include lexical items with varying complexity and would include a wider range of vowel environments. Treated wordlists would be designed in advance and would be recorded at baseline and post-therapy for more accurate measurement of speech outcomes.

This chapter has presented speech outcomes measures based on the perceptual assessment by the SLT, with intra- and inter-rater reliability measures, highlighting areas of strength and weaknesses in perceptual assessment of speech in individuals with CP. The following chapters will present further investigation into the speech data of Andrew and Craig to overcome some of the difficulties highlighted with phonetic transcriptions. Two further methods will be presented. Firstly, a multi-listener perceptual evaluation (reported in Roxburgh et al. 2016) (see chapter 3 and Appendix 10, sub-section 7.10) and secondly, an articulatory analysis of the ultrasound data (see chapter 4).

3 Perception Study

In this section a perceptual evaluation of listener judgements is presented, to supplement some of the issues arising from phonetic transcription and inter-rater reliability. The perceptual evaluation was used to compare Therapy block one (VAM) and Therapy block two (UVBF) presented in the previous section, for both Andrew and Craig. The perceptual evaluation aimed to determine whether an improvement in the target sound, taken from the 36 single-words in the untreated wordlists, was detected by multiple phonetically trained listeners. This section will also present a methodological discussion on whether the perceptual evaluation reported is a potential tool for evaluating therapy outcomes generally.

3.1 Introduction to Perceptual Evaluation

The previous chapter presented therapy outcome measures as PCC and PTCC scores derived from phonetic transcriptions, along with intra- and inter-rater reliability measures. Sell (2005), discusses the issues related to phonetic transcription (see sub-section 0). To circumvent some of the issues with phonetic transcription, perceptual experiments have been used to evaluate post-therapy speech outcomes. Britton et al. (2014) state that perceptual assessment of speech of individuals with CP should include robust listening procedures, use multiple phonetically-trained listeners, and include inter- and intra- rater reliability within this robust procedure. Due to difficulties in phonetic transcriptions for both Andrew and Craig and reliability being lower than the 80% for narrow transcriptions (Shriberg and Lof 1991; Shriberg et al. 1997), this highlighted the need for further perceptual evaluation using a wider range of listeners, for example those without a specialism in CP. This section of the thesis will address the perceptual assessment further, by presenting a multiple phonetically-trained listener perceptual evaluation of Craig and Andrew's data, using robust listening procedures. By presenting the PCC and PTTC scores, along with the perceptual evaluation presented in the current section, this thesis therefore adheres to the requirements for perceptual assessment of the speech of individuals with CP in Britton et al. (2014).

Lohmander and Olsson (2004)'s earlier review of perceptual assessment of the speech of individuals with CP found that many of the studies (28 of 88) used only one listener, with only eight studies using more than 10 listeners. An interval scale was the most common method of judgement, with phonetic transcriptions only being used in eight out of 88 studies. It was concluded that many of the studies did not use or report reliability measures. As reported in Roxburgh et al. (2016), multiple-listener perceptual evaluations of therapy outcomes are more likely to be adopted if they avoid the requirement for high levels of phonetic training, narrow transcription, and cross-transcriber discussion. Castick et al. (2017) presented a comparison of Ordinal Scales, such as that used in CAPS-A (Sell et al. 2009), and Visual Analogue Scales (VAS) (Munson et al. 2012; Baylis et al. 2015). Using expert listeners, they asked participants to rate the speech of individuals with CP using both scales, one month apart. Results showed that both the ordinal scale and VAS were useful tools for evaluating speech in individuals with CP.

An alternative method for perceptual evaluation is a forced-choice judgement between two tokens of recorded speech taken from different stages in treatment, requiring no skills beyond an ability to make a mutual comparison of words, and no specialised knowledge beyond an intuitive grasp of generally agreed norms and the range of variants that are acceptable in the target language. Preferably such a mutual comparison of spoken forms with each other, in reference to the target, should be undertaken independently by multiple listeners (for logistical simplicity) who can focus in detail on a single speaker in order that the listener can become attuned.

The literature suggests that both phonetically trained and untrained listeners should be used for these generic perceptual evaluations and compare results from trained and untrained listeners, where possible (for example, Sell 2005; Brunnegard et al. 2009). It is reported that while specialist SLTs are more reliable than non-specialist SLTs (Keuning et al. 1999), they can often be too critical in their assessment. Non-specialist SLTs, or naïve listeners, offer real-life significance to clinical speech assessments (Sell 2005) and are a useful adjunct to specialist SLT assessment for acceptability measures (Bagnall and David 1988). While using naïve listeners in perceptual evaluations could theoretically add validity, the methodology for this is not yet adequately developed (Sell 2005). The current study only includes phonetically trained listeners. Some listeners had previous experience with assessing children with CLP and others had no experience.

In this section a novel perceptual evaluation is presented, which is intended to be easy to use for both the clinical researchers and the listeners, whether listeners are phonetically-trained or lay listeners. It is intended to supplement the phonetic transcriptions presented in sections 2.2.3 and 2.3.3. Pre- versus post-therapy versions of the 36 single words in the untreated wordlists from both Andrew and Craig were presented, with phonetically trained listeners primed to focus on the speech sounds targeted in therapy, by telling them the therapy targets prior to the experiment (i.e. /n/ for Andrew and velars for Craig). While VAS may allow for near differences in tokens from two different sessions, the design in the current study does not. The goal for therapy is for an improvement in understandability and acceptability, essentially for speech to sound "better" post-therapy. The forced-choice design presented here allows listeners to choose from two tokens taken from two separate sessions (pre- and post-therapy) which one sounds closer to the target, or "better".

Listeners were asked to compare two audio versions of the same word, taken from two different assessment sessions, and choose one as being "closer to the English target", with no specific instructions about the relative importance of phonological, phonetic or prosodic differences between the tokens other than the knowledge of the therapeutic target. Unlike Visual Analogue Scales, or ordinal scales, this comparative method does not require participants to have prior knowledge about what is at either end of a scale or what "mild, moderate or severe" is. By comparing two acoustic tokens, the procedure is intended to be able to discriminate small improvements in speech that is both near-target and hence likely to be "correct" in a PTCC transcription, and in speech that is severely disordered and hence likely to be "incorrect", avoiding some effects of incorrect phonetic transcriptions (Buckingham and Yule 1987). For example, listeners are told that the therapy target for Craig is velars. Where the velar is correct in one version of a word (e.g. "car", [kar]), but incorrect in the other version ([ar]), listeners are likely to select the correct version as "closer to the target".

The results of two perceptual evaluations (sub-study 2a: "non-intervention" and sub-study 2b: "pre/post intervention") designed to investigate listener judgements at various time-points over two blocks of therapy for two children with repaired submucous CP (Andrew and Craig) are presented. The children first received a block of therapy using Speech Trainer 3D followed by a block of UVBF therapy (see sub-section 2.1.9). The design of the perception studies aimed to determine whether listeners detected an improvement in the therapeutic target in untreated words by comparing whole-words from various time points with the hypothesis that words recorded further in the therapy timeline would be closer to the intended target than those recorded pre-therapy.

3.2 Research Questions and Hypotheses

3.2.1 Sub-Study 2a: Non-Intervention Comparisons Research Questions

1. Do listeners select chronologically later in the therapy period (time point B) as "closer to the English target" more often than chronologically earlier in the therapy period (time point A)? Three specific comparisons are considered:
   - BL Comparison: Baseline (assessment 1) and pre-VAM (assessment 2) – one week apart.
   - VAM-UVBF Comparison: post-VAM (assessment 3) and pre-UVBF (assessment 4) – five weeks apart.
   - UVBF-M Comparison: post-UVBF (assessment 5) and maintenance (assessment 6) – three months apart.

Hypothesis: As there is no therapy between the pair of baseline recordings in the no-intervention comparisons, listeners will select equal numbers of A and B within each of the three comparisons. The null hypothesis is that there will be no spontaneous changes between the baselines, and no therapeutically-caused generalisations (or losses) between the two phases or recordings or between the post-therapy maintenance probes.

3.2.2 Sub-Study 2b: Pre/Post Intervention Comparisons Research Questions

1. Do listeners select chronologically later in the therapy period (time point B) as "closer to the English target" more often than chronologically earlier in the therapy period (time point A)? Three specific comparisons are considered:

- VAM (pre/post) Comparison: immediately before and after therapy block one, using VAM
- UVBF (pre/post) Comparison: immediately before and after therapy block two, using U- VBF
- BL-M Comparison: baseline (Assessment session 1 prior to any therapy) to maintenance (Assessment session 6: 3 months after both blocks of therapy)

Hypothesis: Listeners will select later time-points as being “closer to the adult target” if therapy is successful.

2. When listeners select the later time point (post-therapy), are they more confident in their responses than when they select the earlier time point (pre-therapy). The three above comparisons were considered.

Hypothesis: Listeners will be more confident when they select later time-points than earlier time-points.

3. Are listeners reaction times shorter when they select post-therapy as being “closer to the English target” than when they select pre-therapy. The above comparisons were considered.

Hypothesis: Reaction times will be shorter when listeners select post-therapy as being “closer to the English target”.

4. Do lower reaction times correlate with higher levels of confidence in listener responses?

Hypothesis: Words with higher confidence ratings will correlate with shorter reaction times showing that when listeners are more confident in their responses they respond quicker.

3.3 Perceptual Evaluation Method

3.3.1 Participants

3.3.1.1 Speakers

Speakers were Andrew and Craig from the Therapy Study. As discussed in sections 2.2 and 2.3, Andrew was 9;2 years old and was backing /ŋ/ to palatal or velar placement with suspected double articulations. He had previously received extensive therapy to target his production of /ŋ/, with no success. Craig was 6;2 years old and had few high-pressure consonants. He was backing /k/ to glottal placement and fronting /ɡ/ to [d] or [n], with suspected double articulations. These difficulties formed the basis of the therapy discussed in sub-section 2.1.9 and therefore are the core of the materials being evaluated here.

3.3.1.2 Listeners

Listeners were 24 phonetically trained listeners, three male, 21 female, recruited from a university in Central Scotland. All listeners had English as a first language, with mixed Scottish, Irish and English accents. Listeners with known speech, language or hearing impairments were excluded from the study. Five listeners were qualified SLTs working at the university, with the remaining 19 being SLT students who had completed phonetics training as part of the undergraduate and postgraduate programmes. Six listeners had previous experience in working with children with CLP. Experience held by participants was observing patients with CLP on clinical placements and carrying out GOS.SP.ASS'98 assessments.

Listeners were allocated to groups for listening to either Andrew or Craig for sub-study 2b (pre/post intervention). For example, they were allocated to the group for listening to Andrew (Andrew-Group 1 or Andrew-Group 2) for sub-study 2a (no-intervention), they were then allocated to the group for listening to Craig (Craig-Group 1 or Craig-Group 2) for sub-study 2b (pre-post intervention). Counterbalancing was used to reduce listener bias. Listeners' level of experience was not equally spread across both speakers, as students had already been allocated to a group before qualified SLTs were recruited. Table 42 gives an overview of the group allocation for both sub-studies. Twelve listeners evaluated Andrew and 12 listeners evaluated Craig. One listener withdrew from the study and one listener's data was lost. Therefore, 10 listeners evaluated Craig and 12 listeners evaluated Andrew in the no-intervention sub-study and 10 listeners evaluated Andrew, leaving 12 listeners evaluating Craig in the pre-post intervention sub-study.

Participant Number	Sub-Study 2a: Group Allocation	Sub-Study 2b: Group Allocation
1	Andrew - Group 1	Craig - Group 1
2	Andrew - Group 1	Craig - Group 1
3	Andrew - Group 1	Craig - Group 1
4	Andrew - Group 2	Craig - Group 2
5	Andrew - Group 2	Craig - Group 2
6	Andrew - Group 2	Craig - Group 2
7	Craig - Group 1	Andrew - Group 1
8	Craig - Group 1	Andrew - Group 1*
9	Craig - Group 1	Andrew - Group 1
10	Craig - Group 1	Andrew - Group 1
11	Craig - Group 2	Andrew - Group 2
12	Craig - Group 2	Andrew - Group 2
13	Andrew - Group 2	Craig - Group 2
14	Andrew - Group 1	Craig - Group 1
15	Craig - Group 1	Andrew - Group 1
16	Craig - Group 2	Andrew - Group 2
17	Craig - Group 2	Andrew - Group 1**
18	Craig - Group 2	Andrew - Group 1**
19	Craig - Group 2	Andrew - Group 1
20	Andrew - Group 2	Craig - Group 2
21	Andrew - Group 1	Craig - Group 1
22	Andrew - Group 1	Craig - Group 1
23	Craig - Group 1	Andrew - Group 1*
24	Andrew - Group 1	Craig - Group 1

Table 42 Listener Allocations for Sub Study 2a and Sub Study 2b (* indicates where data was missing; ** indicates where participants were reallocated into a different group to account for missing data)

3.3.2 Study Design

3.3.2.1 Therapeutic Design

Each child received six assessment/recording sessions and two blocks of therapy, each with eight one-hour therapy sessions with either VAM or UVBF. Assessments and therapy were carried out by a qualified SLT (the author). See section 2.1 for the general therapy method and sections 2.2 and 2.3 for therapy details for both children.

3.3.2.2 Multiple-Listener Perceptual Evaluation

3.3.2.2.1 Experimental Design

The design of the perceptual evaluation is a modification of a two-alternative forced choice experimental design; using data from the untreated wordlist probes (see sub-sections 2.2.2.2.2 and 2.3.2.2.2). Two sub-studies were designed, both using the same experimental design (sub-study 2a: “non-intervention” and sub-study 2b: “pre/post intervention”) with three comparisons per sub-study. Comparisons were between pairs of tokens of the same word drawn from two different time points during therapy (assessment sessions one to six). There were no comparisons of two tokens taken from the same session. Listeners were told they would hear two versions (first V1 then V2) of the same single real target word. The order of presentation was counterbalanced, so that either V1 or V2 could be earlier (A) or later (B) in therapy. Listeners were asked to decide which acoustic stimulus sounded “closer to the English target word” presented orthographically on the screen.

3.3.2.2.2 Exporting the Audio Data

Speech materials were 36 single-words taken from speaker-specific untreated wordlists for both Andrew (/n/) and Craig (velars). Audio data was exported from AAA into PRAAT version 5.3.57 (Boersma and Weenink 2013). Individual words were edited from longer recordings (three words per recording) hence silence was included either side of each word where possible (Figure 35). Single words were saved as individual WAV. files for creating the multiple forced choice (MFC) script (see below). An additional 0.5 second silence was presented between V1 and V2 as the inter-stimulus interval. Volume of audio data was not controlled during recordings, therefore audio data from each session was at a different loudness level. The number of tokens for each of the three comparison blocks was 36 single words, giving a total of 108 comparisons for each sub-study.

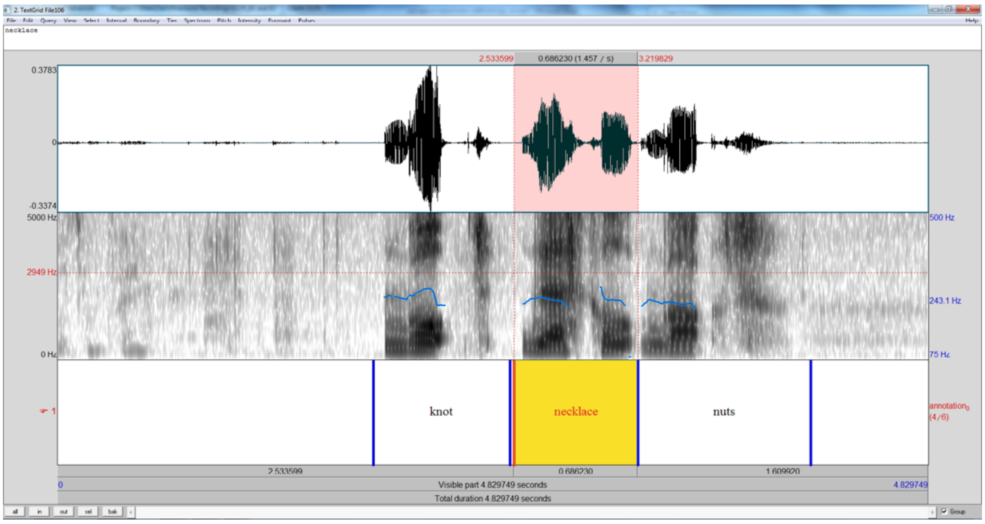


Figure 35 Example of cropping audio data in PRAAT to create individual WAV. Files

3.3.2.2.3 Designing the Experiments in PRAAT

The experiment was designed by modifying a previously existing Multiple Forced Choice (MFC) experiment in PRAAT version 5.3.57 (Boersma and Weenink 2013). The order of tokens was counterbalanced, so that half of the tokens, from the 36 untreated words (per speaker) were presented in AB (earlier/later) order to half of the listeners and BA (later/earlier) to the other half of the listeners. The number of tokens for each comparison was 36 single words, giving a total of 108 comparisons for each sub-study, with no repetitions of the same word from the same assessment session. As listeners could listen to each comparison up to three times, as suggested by Shriberg and Kwiatkowski (1980) as the maximum time listeners should hear each token, this meant that they were listening to a maximum of 324 word pairs for each sub-study. The number of times each word was heard was controlled and the order of the words was randomised for each listener by adding in <PermuteBalanceNoDoublets> (i.e. randomisation of tokens with no duplicates) into the MFC document. A self-determined break was provided every 18 comparisons (one break per comparison). A copy of the MFC script can be found in Appendix 9 (sub-section 7.9).

3.3.2.2.4 Running the Experiments in PRAAT

The experiment took place in a laboratory in a university over two sessions. Sub-study 2b took place approximately one month after sub-study 2a. Listeners were asked to compare two audio versions of the same word, taken from two different assessment sessions, and chose one as being “closer to the English target”. Figure 36 shows the PRAAT interface presented to listeners.

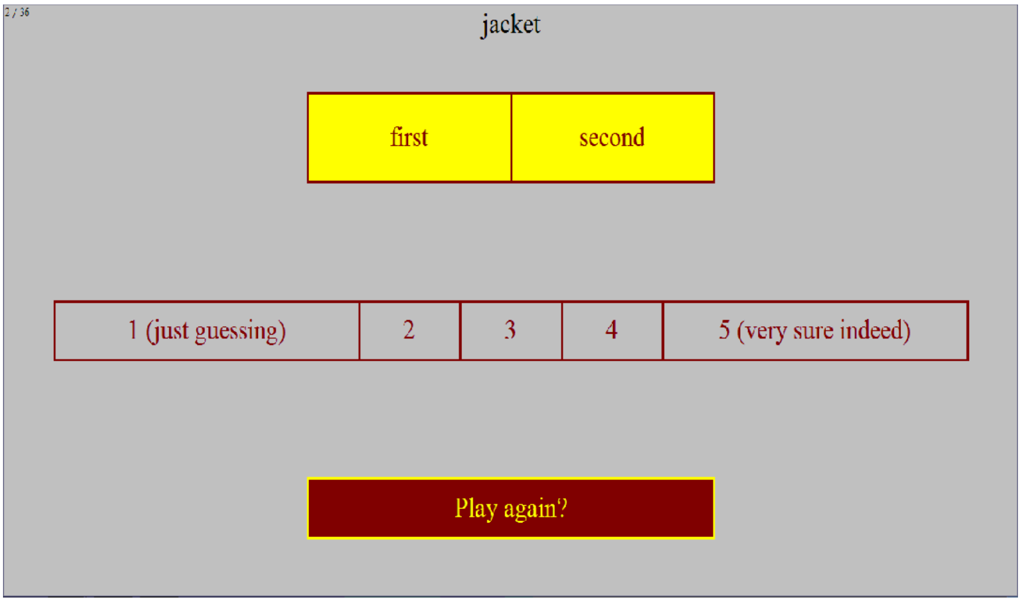


Figure 36 PRAAT interface presented to listeners

Data was presented to listeners using PRAAT version 5.3.57 (Boersma and Weenink 2013). Listeners were placed at individual computers with personal headphones in order for them to control the experiment themselves and listen to the data at their own pace. During the no-intervention comparisons (sub-study 2a), listeners were asked to complete three individual evaluations for each of the comparisons. In the pre/post intervention comparisons (sub-study 2b), all three comparisons were merged into one experiment to increase randomisation of tokens. They were told they could listen to each token three times. They were asked to select the version they believed to be more target-like and to rate their confidence on a scale of one-to-five, with one being least confident and five being most confident in their response. The software automatically recorded the time taken for listeners to respond, calculated from when the listeners first heard version one to when they clicked their response, in order to calculate reaction times.

3.3.2.2.5 Exporting the Data from PRAAT

Once each participant had completed the experiment, the researcher extrapolated data from PRAAT and exported it for analysis.

3.3.3 Data Analysis

In order to answer research question 1 from both sub-study 2a (no-intervention) and sub-study 2b (pre/post-intervention), a non-parametric Sign test (Corder and Foreman 2014) was used for statistical analysis of each of a listener's three blocks of comparisons in each sub-study. The number of words from session B (later) judged to be more target-like for each individual listener was tested for significance at $p<.05$. For a two-tailed test of 36 lexical pairs, this requires a listener to assign more than 25 to one or other category ($p=.029$).

Since it can be argued that the three comparison blocks in each sub-study are not independent, significance will also be presented at $p/3$ as a Bonferroni adjustment, i.e. significance is set at $p<.017$ (a threshold of 26/36, $p=.011$). Listener agreement was then calculated, based on the overall number of session A or session B selected (including non-significant preferences) and was statistically tested using a Fleiss' Kappa (Fleiss 1981) for each word-pair comparison. A Fleiss' Kappa was also calculated for WI, WM and WF contexts. A Fleiss' Kappa can be used to measure agreement among listeners and transcribers (see below). Fleiss' Kappa results can be interpreted in the following way: $< .40$ = Poor agreement; $.60 – .74$ = Intermediate to good agreement; $\geq .75$ = Excellent agreement (Fleiss 1981).

A paired t-test was used to measure the difference in confidence levels when listeners chose A (earlier/pre-therapy) or B (later/post-therapy). A paired t-test was also used to measure the difference in reaction times when listeners chose A or B.

A Pearson's Correlation was used to measure the relationship between confidence in listener responses and reaction times. A negative correlation would be expected, showing listeners respond quicker when they are more confident in responses. Two correlations were made for each comparison (VAM, UVBF and BL-M) separating A and B responses.

3.4 Perceptual Evaluation Results

3.4.1 Andrew: No-Intervention Comparison Sub-Study

3.4.1.1 Listener Responses

In the BL-VAM Comparison, to test for baseline stability pre-therapy, listeners selected the later time-point, i.e. “B”, for 267/432 comparisons (62%). In other words, 67% of Andrew’s productions in the pre-VAM session were judged as closer to the target, suggesting that the two sessions pre-therapy (BL: assessment sessions 1 and pre-VAM: assessment session 2) were not stable. Rising baselines are not recommended for single-subject research. However, PTCC scores indicated stability between the baseline and pre-VAM sessions (see sub-section 2.2.3.4). As this perceptual evaluation was not carried out until after the therapy study, PTCC scores were used for measuring stability in the baseline phase and deciding on suitability for therapy. Previous studies, such as Cleland et al. (2015c), also use PTCC and do not use multi-listener perceptual evaluations to measure baseline stability.

In the VAM-UVBF Comparison, later (B) was selected for 224/432 tokens (52%), showing the pre-UVBF productions were selected more than post-VAM session. In the UVBF-M Comparison 258/432 (60%) of B were selected, showing that post-UVBF therapy productions were selected more than the maintenance. Table 43 and Figure 37 provide a summary of the pooled data.

	BL-VAM Comparison	VAM-UVBF Comparison	UVBF-M Comparison
Tokens	267 (62%)	224 (52%)	258 (60%)
Listener trend	11	6	12
Listener significance	2	0	0

Table 43 Numbers of more adult-like judgements for later session of the three comparisons undertaken, all data pooled for Andrew in the no-intervention sub-study and the number of listeners for whom a statistically significant individual judgement was indicated. A negative listener number indicates a preference for the earlier session.

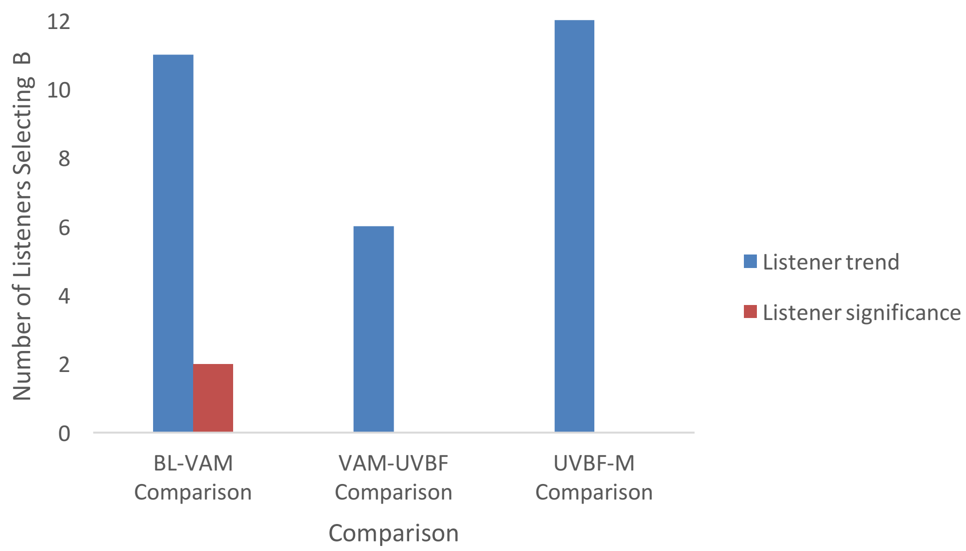


Figure 37 Pooled data for Andrew Sub-Study 2a

Statistical analysis showed that within the BL-VAM Comparison, 11 listeners selected B (later in time) as closer to target in raw proportion, but only three listeners selected B statistically significantly more than A (Listener 5, *p* =.0012; Listener 13, *p*=. 0012; Listener 20, *p*=.0288). When Bonferroni adjusted, only two listener's responses were significant (Listener 5 and Listener 13). One listener selected A and B equally (Listener 14). Within the VAM-VBF Comparison, six listeners selected B (later in time) as closer to target more often than A with only one listener (Listener 2) selecting A more than B, although not significant (presented in grey in Table 44). Four listeners selected A and B equally within the VAM-UVBF Comparison. In the UVBF-M Comparison, all listener selected B more than A, with only one listener (Listener 13) producing significant results (*p* = .0288). However, when Bonferroni adjusted Listener 13's results were in fact not significant. Results for each individual listener are outlined in Table 44, with significance denoted by boldface.

Listener Number	BL-VAM Comparison	VAM-UVBF Comparison	UVBF-M Comparison
1	.4050	1.1321	.0652
2	.2430	.0652	.4050
3	.4050	.8679	.6177
4	.8679	.8679	.1325
5	**.0012****	.0652	.6177
6	.1325	1.1321	.6177
13	**.0012****	1.1321	**.0288***
14	1.1321	.2430	1.1321
20	**.0288***	.4050	.1325
21	.8679	1.1321	.1325
22	.2430	.8679	.1325

Table 44 Sign Results for Andrew for all three comparisons in the no-intervention sub-study (*p < .05 **p < .01 ***p <.001)

3.4.1.2 Listener Agreement: Word-Level Analysis

Results show that 100% listener agreement was found in 8/36 words in only one comparison per word. In other words, there were no words that had 100% agreement in more than one comparison. Fleiss' Kappa results (Table 45) show that there was general agreement between listeners for all word positions in all three comparisons, with the lowest agreement found in WM position in the BL-VAM Comparison (Fleiss' Kappa = 0.0733) and the highest agreement found in WI position in the BL-VAM Comparison (Fleiss' Kappa = 0.3501).

Comparison	Listener Agreement: Fleiss' Kappa, 95% CI
BL-VAM: Word Initial	0.3501 (0.2805, 0.4198)
BL-VAM: Word Medial	0.0733 (0.0037, 0.1430)
BL-VAM: Word Final	0.3348 (0.2652. 0.4045)
VAM-UVBF: Word Initial	0.1966 (0.1270, 0.2663)
VAM-UVBF: Word Medial	0.3439 (0.2742, 0.4135)
VAM-UVBF: Word Final	0.1540 (0.0843, 0.2236)
UVBF-M: Word Initial	0.1307 (0.0610, 0.2003)
UVBF-M: Word Medial	0.1959 (0.1262, 0.2655)
UVBF-M: Word Final	0.1919 (0.1222, 0.2615)

Table 45 Andrew Fleiss' Kappa results for word positions in each comparison in sub-study 2a: no-intervention

3.4.2 Andrew: Pre/Post Intervention Sub-Study

3.4.2.1 Listener Responses

Overall, in the BL-M Comparison, listeners selected B for 254/360 comparisons (71%). In other words, 71% of Andrew's productions in the maintenance session were judged as closer to the target, suggesting improvement. In the VAM Comparison, B was selected for 228/360 tokens (63%), showing the post-therapy productions were selected more than pre-therapy, and in the UVBF Comparison, 151/360 (42%), unexpectedly showing that pre-therapy productions were selected more than post-therapy productions. Table 46 and Figure 38 provide a summary of the pooled data.

	VAM Comparison	UVBF Comparison	BL-M Comparison
Tokens	228 (63%)	151 (42%)	254 (71%)
Listener trend	10	3	10
Listener significance	2	-1	6

Table 46 Numbers of more adult-like judgements for later session of the three comparisons undertaken, all data pooled for Andrew in the pre/post intervention sub-study, and the number of listeners for whom a statistically significant individual judgement was indicated. A negative listener number indicates a preference for the earlier session.

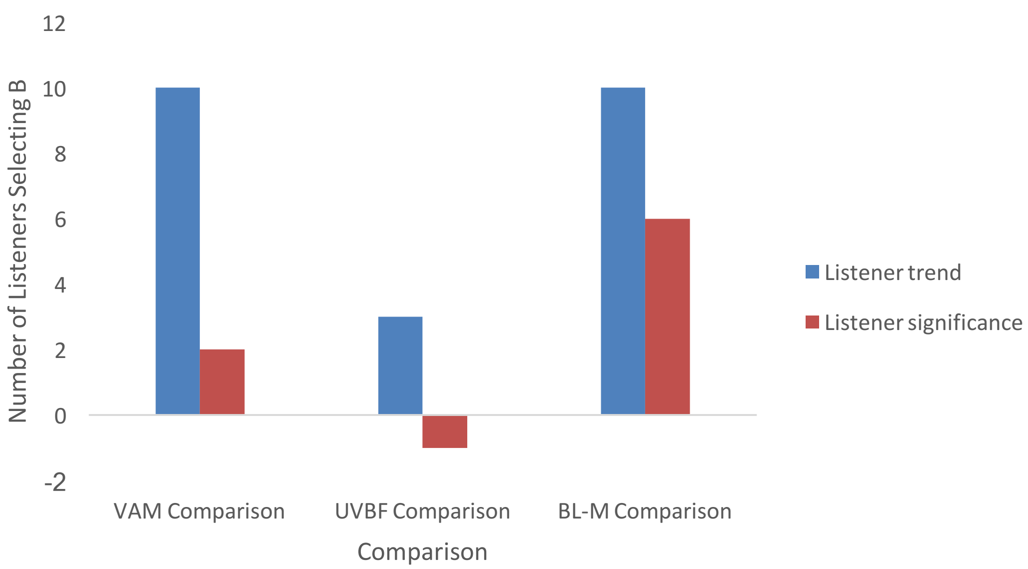


Figure 38 Pooled Data for Andrew Sub-Study 2b

Statistical analysis showed that within the VAM Comparison, all listeners selected B (post-therapy) as closer to the target but only two listeners selected B statistically significantly more than A (Listener 17, p=.0039; Listener 11, p=. 0113, both still significant at the adjusted threshold of p<.017). Within the UVBF Comparison, a very different trend was observed, in which 7/10 listeners selected A (pre-therapy) as closer to target more often than B (denoted by grey text in Table 47), though only one (Listener 18) selected A significantly more than B (p <.0001). In the BL-M Comparison, all listeners selected B more than A, with 7/10 listeners producing significant results, or 6/10 with Bonferroni adjustment. Results for each individual listener are outlined in Table 47, with significance denoted by boldface. Overall, listeners selected B more often than A, but more tokens of A were selected in the UVBF Comparison, showing that although there was an overall improvement after both types of therapy, there was statistically no change, and perhaps suggesting deterioration post-therapy in the UVBF comparison.

Listener Number	**VAM Comparison**	**UVBF Comparison**	**BL-M Comparison**
7	.4050	.2430	**.0012****
9	.2430	.4050	.1325
10	.6177	.8679	**.0012****
11	**.0113****	.1325	.0652
12	.0652	.4050	**.0288****
15	.0652	.4050	**.0113****
16	.1325	.6177	**.0113****
17	**.0039****	.6177	**.0113****
18	.4050	**<.0001*****	.2430
19	.6177	1.1321	**.0113****

Table 47 Sign Results for Andrew for all three comparisons (*p < .05 **p < .01 ***p <.001)

3.4.2.2 Listener Agreement: Word-Level Analysis

Results show that 100% listener agreement was found in 18/36 words in at least one comparison. Fleiss' Kappa results show that there was general agreement between listeners for all word positions in all three comparisons, with the lowest agreement found in WI position in the VAM Comparison (Fleiss' Kappa =.0770) and the highest agreement found in WF position in the UVBF Comparison (Fleiss' Kappa =.4351). Both of these results demonstrate poor agreement between listeners, based on Fleiss (1981) interpretation. Table 48 shows Fleiss' Kappa results for each word positions within each comparison.

Comparison	Listener Agreement: Fleiss' Kappa, 95% CI
VAM: Word Initial	0.0770 (0.0074, 0.1613)
VAM: Word Medial	0.2361 (0.1518, 0.3205)
VAM: Word Final	0.2002 (0.1159, 0.2846)
UVBF: Word Initial	0.3131 (0.2288, 0.3974)
UVBF: Word Medial	0.2603 (0.11760, 0.3447)
UVBF: Word Final	0.4351 (0.3507, 0.5194)
BL-M: Word Initial	0.2395 (0.1552, 0.33239)
BL-M: Word Medial	0.3290 (0.2447, 0.4134)
BL-M: Word Final	0.2157 (0.1313, 0.3000)

Table 48 Andrew Fleiss' Kappa results for word positions in each comparison for sub-study 2b: pre/post intervention

3.4.2.3 Confidence Levels and Reaction Times

VAM Comparison: Confidence ratings (1=least confident, 5=most confident) were compared between pre-therapy responses (*M*=2.21, *SD*=0.87) and post-therapy responses (*M*=2.71, *SD*=0.86) for the VAM comparison. A paired t-test showed no significant difference in listener's confidence in their responses if they selected pre- (A) or post-therapy (B) (t=0.06). Reaction time (seconds) in listeners' responses were also compared for when they selected pre-therapy (*M*=3.77, *SD*=0.77) and post-therapy (*M*=3.55, *SD*=0.58). A paired t-test showed no significant difference between listener's confidence in pre- and post-therapy selection (t=0.11).

A Pearson's correlation was used to analyse the relationship between reaction times and listener's confidence levels. When listeners selected post-therapy tokens, no correlation was found (r=-0.161, p=0.363). When listeners selected pre-therapy tokens, a statistically significant weak negative correlation was found (r=-0.499, p=0.003**). Figure 39 shows the scatter plot for the correlation between listener confidence and reaction times in the VAM comparison. Although it shows that there is a trend between lower reaction times and higher levels of confidence, statistical analysis shows that this relationship is weak.

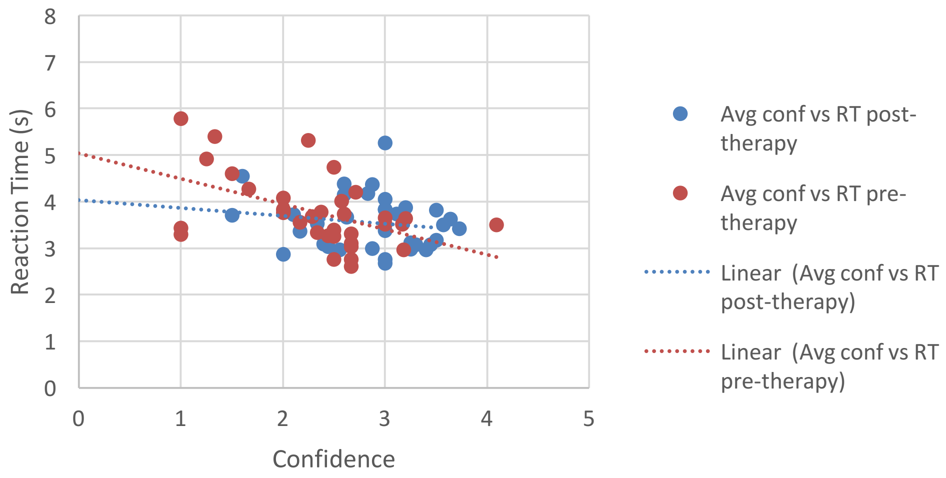


Figure 39 Andrew Correlation of reaction time and confidence levels for the VAM comparison (Andrew)

UVBF Comparison: Confidence ratings were compared between pre-therapy responses (*M*=3.19, *SD*=0.64) and post-therapy responses (*M*=2.15, *SD*=1.11) for the UVBF comparison. A paired t-test showed a statistically significant difference in listener's confidence in their responses if they selected pre- or post-therapy (*t*=0.00**). Reaction time in listeners' responses were also compared for when they selected pre-therapy (*M*=3.64, *SD*=0.59) and post-therapy (*M*=3.71, *SD*=0.55). A t-test showed no significant difference between listener's confidence in pre- and post-therapy selection (*t*=0.92).

A Pearson's correlation showed that when listeners selected post-therapy tokens, no correlation was found (*r*=-0.330, *p*=0.070). When listeners selected pre-therapy tokens, no correlation was found (*r*=-0.031, *p*=0.861). Figure 40 shows the scatter plot for the correlation between listener confidence and reaction times in the UVBF comparison. Although it suggests a trend between lower reaction times and higher levels of confidence, like the VAM comparison, statistical analysis shows that there is no real relationship.

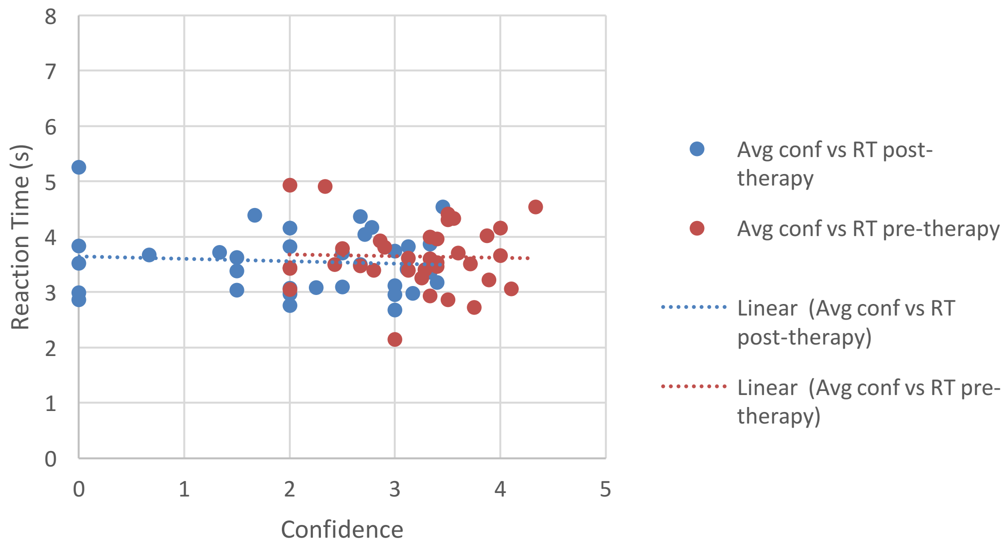


Figure 40 Correlation of reaction time and confidence levels for the UVBF comparison (Andrew)

BL-M Comparison: Confidence ratings were compared between pre-therapy (*M*=3.21, *SD*=0.81) responses and post-therapy responses (*M*=3.41, *SD*=0.71) for the BL-M comparison. A paired t-test showed no significant difference in listener's confidence in their responses if they selected pre- or post-therapy (*t*=0.85). Reaction time in listeners' responses were also compared for when they selected pre-therapy (*M*=3.61, *SD*=0.77) and post-therapy (*M*=3.45, *SD*=0.54). A t-test showed no significant difference between listener's confidence in pre- and post-therapy selection (*t*=0.50).

A Pearson's correlation was used to analyse the relationship between reaction times and listener's confidence levels and none were found. Figure 41 shows the scatter plot. Similar to the VAM and UVBF comparisons, there is no significant relationship.

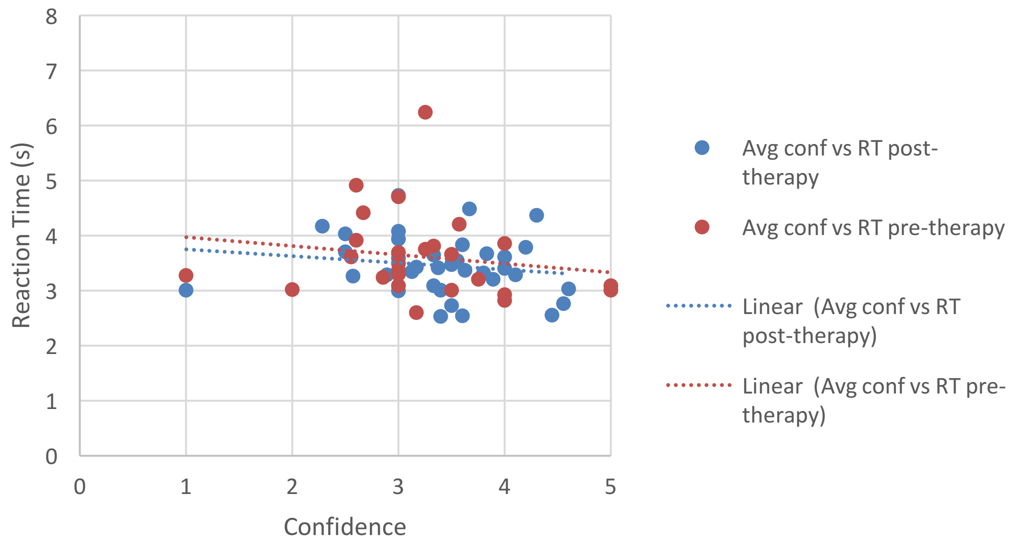


Figure 41 Correlation of reaction time and confidence level for the BL-M comparison (Andrew)

3.4.3 Andrew: Results Summary

For sub-study 2a (no-intervention comparisons), listeners selected B (later in time) more often than A (earlier in time) within all three comparisons (BL-VAM, VAM-UVBF and UVBF-maint), with only one listener selecting B significantly more than A in the VAM-UVBF comparison. For sub-study 2b (pre/post therapy), listeners also selected B (post-therapy) more than A (pre-therapy) for the VAM and BL-M comparisons, however contrary to expectations, listeners selected A more than B in the UVBF comparison. This corresponds with the mean PTCC scores from transcribers, with higher PTCC scores found in the same sessions selected as being "closer to the English target". In the UVBF Comparison, when listeners selected A more than B, PTCC also decreased in the UVBF post-therapy session, therefore both measures indicate a slight deterioration in this block of therapy. Data from inter-rater reliability measures (see sub-section 2.2.3.4.3) showed only 59% agreement in the UVBF comparison, suggesting that the data was more difficult to interpret than in the other sessions.

For the VAM and BL-M comparisons, listeners were more confident in their responses and their reaction times were faster when they selected post-therapy versions of the word (B) as being closer to the target, however statistical analysis found no significant differences between whether listeners selected A or B. In the UVBF comparison, listeners were more confident and quicker to respond when they selected A (pre-therapy). A weak negative correlation (quicker reaction times with increased confidence) was found when listeners selected pre-VAM in the VAM comparison. No correlations were found in the UVBF or BL-M comparisons.

3.4.4 Craig: No-Intervention Comparisons Sub-Study

3.4.4.1 Listener Responses

The number of B (later in time) selected within each comparison was calculated. Within the BL-VAM Comparison, listeners selected B for 248/432 tokens (57%) suggesting that Craig's production of velars in single words in Assessment session 2 (pre-VAM) were closer to the target. In the VAM-UVBF Comparison, B was selected for 157/432 tokens (36%), i.e. pre-UVBF recording were selected only 36% of the time, with A being selected more than B. In the UVBF-M Comparison, 236 /432 (55%) showing that productions from the maintenance session were selected 55% of the time. Table 49 and Figure 42 provide a summary of the pooled data.

	BL-VAM Comparison	VAM-UVBF Comparison	UVBF-M Comparison
Tokens	248 (57%)	157 (36%)	236 (55%)
Listener trend	12	1	12
Listener significance	4	-3	0

Table 49 Numbers of more adult-like judgements for later session of the three comparisons undertaken, all data pooled for Craig in the no-intervention Sub-Study, and the number of listeners for whom a statistically significant individual judgement was indicated. A negative listener number indicates a preference for the earlier session.

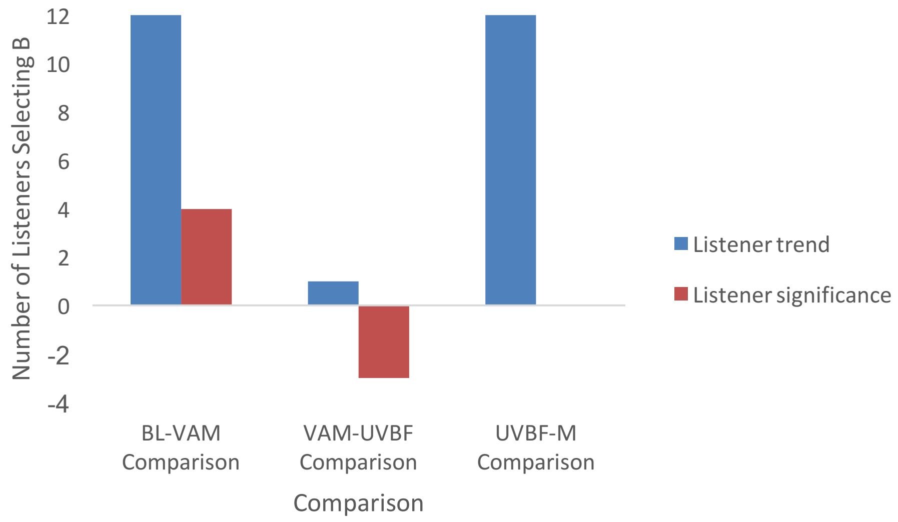


Figure 42 Pooled Data for craig Sub-Study 2a

Statistical analysis of the listener responses using a two-tailed Sign test, with a chance level of $p < .05$ was calculated. In the BL-VAM Comparison, 10 listeners selected B more than A, with two listeners selecting an equal number of A and B. Statistical analysis showed that only four of these preferences for the later session were statistically significant (all at $p < .01$). In the VAM-UVBF Comparison, one listener (Listener 11) selected B and A equally, with 11 listeners selecting A more than B (denoted in Table 50 by grey text). Out of these 11 listeners, four selected A significantly more than B ($p < .05$). Contrary to the expectations that there would be an improvement between these two sessions, 11 listeners selected the earlier time-point, just as strongly suggesting deterioration in Craig's production of velars between the two therapy blocks. In the UVBF-M Comparison, nine listeners selected B more than A, however no individual results were significant. In addition, three listeners in the UVBF-M Comparison selected A and B equally. As the majority of listeners (9/11) selected B more than A, this suggests continuing generalisation between the post-UVBF and maintenance sessions. The results from each of the three comparisons did not change when Bonferroni adjusted. Table 50 shows individual listener results, with significance denoted by boldface.

Listener Number	BL-VAM Comparison	VAM-UVBF Comparison	UVBF-M Comparison
7	.6177	.4050	.8679
8	**.0113****	.2430	.6177
9	.0652	**.0288***	.8679
10	**.0113****	.0652	.4050
11	.8679	1.1321	.4050
12	.6177	**.0288***	.1325
15	**.0113****	.4050	.4050
16	.8679	**.0288***	1.1321
17	1.1321	.2430	1.1321
18	1.1321	**.0288***	1.1321
19	.6177	.2430	.0288
23	**.0113****	.1325	.8679

Table 50 Sign Results for Craig in the no intervention comparisons (NB *p < .05 **p < .01 ***p <.001)

3.4.4.2 Listener Agreement: Word-Level Analysis

Listener agreement was calculated for each word pair within each comparison (BL-VAM, VAM-UVBF and UVBF-M), also with Kappa. Results show that 100% listener agreement was found in 11/36 words in at least one comparison. Fleiss' Kappa results show that there was agreement between listeners for all word positions in all three comparisons, with the lowest agreement found in WI position in the BL-VAM Comparison (Fleiss' Kappa = .0697) and the highest agreement found in WI position in the VAM-UVBF Comparison (Fleiss' Kappa = .3538). Based on Fleiss' (1981) interpretation of results however, both of these results demonstrate poor agreement between listeners on a word-by-word basis. Table 51 shows Fleiss' Kappa results for all word positions in each of the three comparisons.

Comparison	Listener Agreement: Fleiss' Kappa, 95% CI
BL-VAM: Word Initial	0.0697(0.0001, 0.1394)
BL-VAM: Word Medial	0.1202 (0.0505, 0.1898)
BL-VAM: Word Final	0.1742 (0.1046, 0.2439)
VAM-UVBF: Word Initial	0.3538 (0.2842, 0.4235)
VAM-UVBF: Word Medial	0.3199 (0.2502, 0.3895)
VAM-UVBF: Word Final	0.1530 (0.0833, 0.2226)
UVBF-M: Word Initial	0.2899 (0.2202, 0.3595)
UVBF-M: Word Medial	0.3202 (0.2506, 0.3899)
UVBF-M: Word Final	0.1896 (0.1199, 0.2592)

Table 51 Craig Fleiss' Kappa results for word positions in each comparison of sub-study 2a: no intervention

3.4.5 Craig: Pre/Post Intervention Comparisons Sub-Study

3.4.5.1 Listener Responses

The number of B (later in time) selected within each comparison was calculated. Overall, in the VAM Comparison, listeners selected B for 326/432 tokens (75%) suggesting that Craig's production of velars in single words post-therapy were closer to the target 75% of the time. In the UVBF Comparison, B was selected for 249/ 432 tokens (58%), i.e. post-therapy recording were selected 58% of the time, and in the BL-M Comparison, 350 /432 (81%) showing that productions from the maintenance session were selected 81% of the time. Table 52 and Figure 43 provide a summary of the pooled data.

	VAM Comparison	UVBF Comparison	BL-M Comparison
Tokens	326 (75%)	249 (58%)	350 (81%)
Listener trend	12	10	12
Listener significance	11	1	12

Table 52 Numbers of more adult-like judgements for later session of the three comparisons undertaken, all data pooled for Craig in the pre/post intervention comparisons, and the number of listeners for whom a statistically significant individual judgement was indicated. A negative listener number indicates a preference for the earlier session.

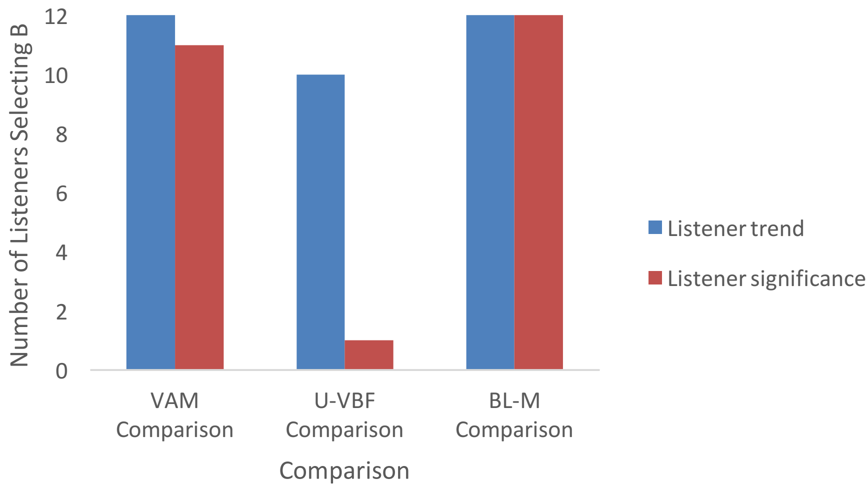


Figure 43 Pooled Data for Craig Sub-Study 2b

Statistical analysis of the listener responses using a two-tailed sign test, with a chance level of $p < .05$ was calculated. Results showed that almost all listeners selected B significantly more than A within the VAM Comparison (11/12, or 10/12 when Bonferroni adjusted). In the UVBF Comparison, only one listener (Listener 20) selected B significantly more than A (and in fact at $p = .0288$, L20 was not significant when Bonferroni adjusted). Two listeners (14 and 13) selected A more than B (denoted in Table 53 by boldface), however neither were significant. Overall there was significant improvement in the overall BL-M Comparison, with all listeners selecting B significantly more than A, even after Bonferroni adjustment. Table 53 shows individual listener results, with significance denoted by boldface.

Listener Number	VAM Comparison	UVBF Comparison	BL-M Comparison
1	**.0003****	1.1321	**<.0001*****
2	**.0003****	.4050	**<.0001*****
3	**.0003****	.8679	**<.0001*****
4	.1325	.8679	**.0113****
5	**.0039****	.1325	**.0003****
6	**.0288***	.1325	**.0003****
13	**<.0001*****	.8679	**.0003****
14	**.0039****	.6177	**<0.0001*****
20	**.0113****	**.0288***	**.0113****
21	**.0113****	.6177	**.0003****
22	**.0003****	.0652	**.0012****
25	**.0113****	.0652	**<0001*****

Table 53 Sign Results for Craig in the pre/post intervention sub-study (NB *p < .05 **p < .01 ***p <.001)

3.4.5.2 Listener Agreement: Word-Level Analysis

Listener agreement was calculated for each word pair within each comparison (VAM, UVBF and BL- M), also with Kappa. Results show that 100% listener agreement was found in 27/36 words in at least one comparison. Fleiss' Kappa results show that there was agreement between listeners for all word positions in all three comparisons, with the lowest agreement found in WI position in the VAM Comparison (Fleiss' Kappa = .0411) and the highest agreement found in WF position in the BL-M Comparison (Fleiss' Kappa = .5927). Based on Fleiss' (1981) interpretation of results however, both of these results demonstrate poor agreement between listeners on a word-by-word basis. Table 54 shows Fleiss' Kappa results for all word positions in each of the three comparisons.

Comparison	Listener Agreement: Fleiss' Kappa, 95% CI
VAM: Word Initial	0.0411 (0.0286, 0.1107)
VAM: Word Medial	0.4508 (0.3811, 0.5204)
VAM: Word Final	0.5121 (0.4425, 0.5818)
UVBF: Word Initial	0.4317 (0.3621, 0.5014)
UVBF: Word Medial	0.4343 (0.3647, 0.5040)
UVBF: Word Final	0.1919 (0.1223, 0.2616)
BL-M: Word Initial	0.4116 (0.3419, 0.4812)
BL-M: Word Medial	0.3368 (0.2671, 0.4064)
BL-M: Word Final	0.5927 (0.5230, 0.6623)

Table 54 Craig Fleiss' Kappa results for word positions in each comparison of sub-study 2b: pre/post intervention

3.4.5.3 Reaction Times and Confidence Levels

VAM Comparison: Confidence ratings (1=least confident, 5=most confident) were compared between pre-therapy responses (A) (*M*=2.69, *SD*=1.07) and post-therapy responses (B) (*M*=3.53, *SD*=1.07) for the VAM comparison. A paired t-test showed no significant difference in listener's confidence in their responses if they selected pre- or post-therapy (*t*=0.26). Reaction time (seconds) in listeners' responses were also compared for when they selected pre-therapy (*M*=4.10, *SD*=1.09) and post-therapy (*M*=3.87, *SD*=0.91). A paired t-test showed no significant difference between listener's confidence in pre- and post-therapy selection (*t*=0.26).

A Pearson's correlation was used to analyse the relationship between reaction times and listener's confidence levels. When listeners selected post-therapy tokens, no correlation was found (*r*=-0.422, *p*=0.016*). When listeners selected pre-therapy tokens, no correlation (*r*=-0.184, *p*=0.412). Figure 44 shows the scatter plot for the correlation between listener confidence and reaction times in the VAM comparison.

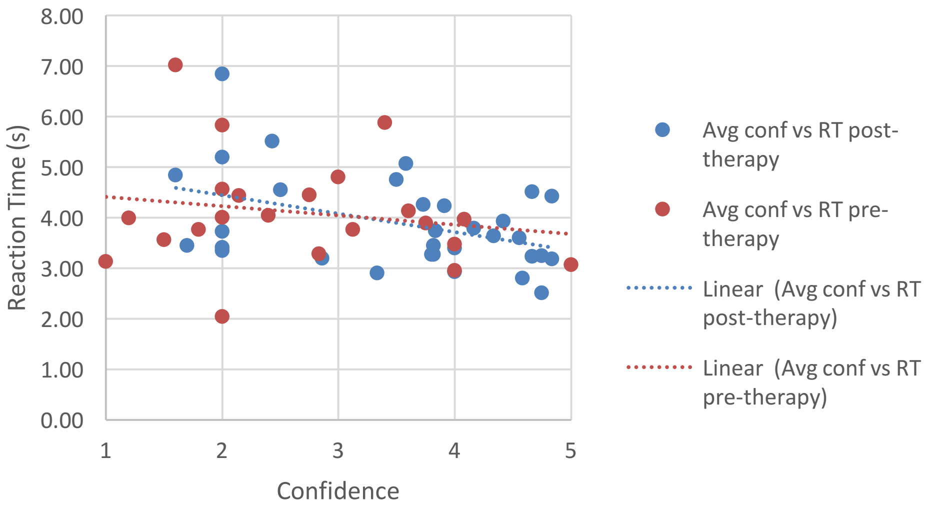


Figure 44 Correlation of reaction time and confidence levels for the VAM comparison (Craig)

UVBF Comparison: Confidence ratings were compared between pre-therapy responses (*M*=2.84, *SD*=0.72) and post-therapy responses (*M*=3.21, *SD*=0.95) for the UVBF comparison. A paired t-test showed no significant difference in confidence in listeners' responses if they selected pre- or post-therapy (*t*=0.28). Reaction time in listeners' responses were also compared for when they selected pre-therapy (*M*=4.47, *SD*=0.93) and post-therapy (*M*=4.44, *SD*=0.94). A paired t-test showed no significant difference between listener's confidence in pre- and post-therapy selection (*t*=0.70).

A Pearson's correlation showed that when listeners selected post-therapy tokens, a statistically significant weak negative correlation was found (*r*=-0.366, *p*=0.036*). When listeners selected pre-therapy tokens, no correlation was found (*r*=-0.105, *p*=0.581), indicating no relationship between lower reaction times and higher levels of confidence. Figure 45 shows the scatter plot for the correlation between listener confidence and reaction times in the UVBF comparison.

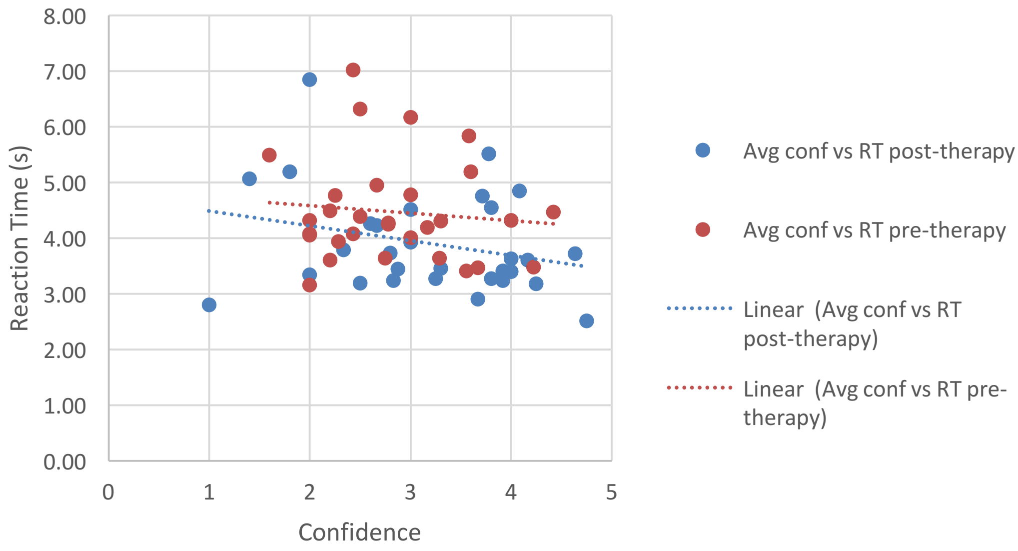


Figure 45 Correlation of reaction time and confidence levels for the UVBF comparison (Craig)

BL-M Comparison: Confidence ratings were compared between pre-therapy responses (*M*=2.88, *SD*=0.75) and post-therapy responses (*M*=3.98, *SD*=0.93) for the BL-M comparison, again suggesting that listeners felt more confident when they selected B (post-therapy). However, a paired t-test showed no significant difference in listener's confidence in their responses (*t*=0.15). Reaction time in listeners' responses were also compared for when they selected pre-therapy (*M*=4.10, *SD*=0.81) and post-therapy (*M*=3.83, *SD*=0.79). A paired t-test showed no significant difference between listener's confidence in pre- and post-therapy selection (*t*=0.71). When listeners selected post-therapy tokens, a statistically significant weak negative correlation was found (*r*=-0.493, *p*=0.003**). Also, when listeners selected pre-therapy tokens, a statistically significant weak negative correlation was found (*r*=-0.516, *p*=0.034*). These both indicate a weak relationship between lower reaction times and higher levels of confidence. Figure 46 shows the scatter plot for the correlation between listener confidence and reaction times. Similar to the VAM and UVBF comparisons, it shows that there is a trend between lower reaction times and higher levels of confidence, however statistical analysis shows that this relationship is weak.

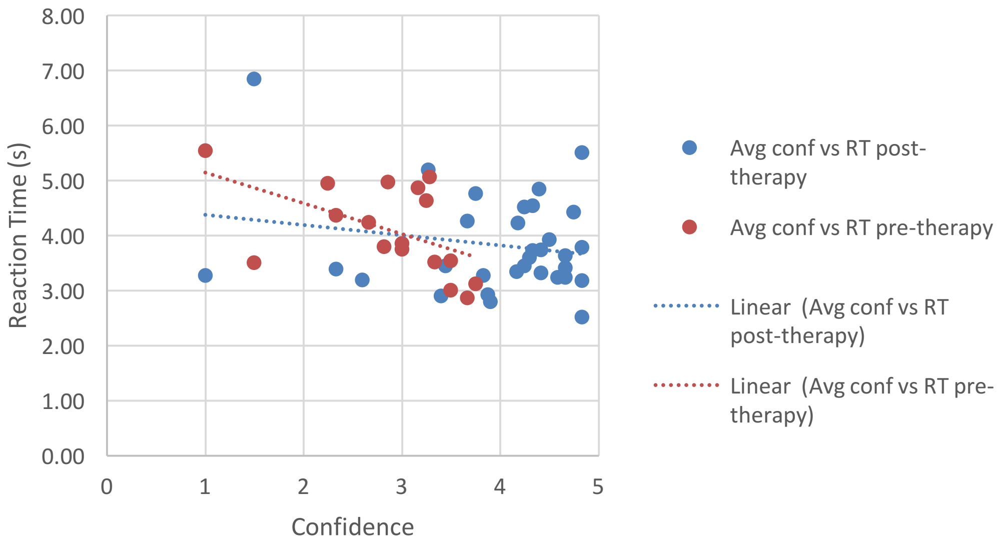


Figure 46 Correlation of reaction time and confidence levels for the BL-M comparison (Craig)

3.4.6 Craig: Results Summary

For sub-study 2a (no-intervention comparisons), listeners selected B (later in time) more than A (earlier in time) within the BL-VAM and UVBF-maint comparisons. Within the VAM-UVBF comparison, listeners selected B (earlier in time) more than A. For sub-study 2b (pre/post therapy), listeners also selected B (post-therapy) more than A (pre-therapy) within all three comparisons (VAM, UVBF and BL-M).

This corresponds with the mean PTCC scores from transcribers, with higher PTCC scores found in the same sessions selected as being closer to the target. All three transcribers from the inter-rater reliability (see sub-section 2.3.3.4.3) reported an increase in PTCC in the VAM comparison (pre-VAM 24% PTCC to post-VAM 84% PTCC), the UVBF comparison (pre-UVBF 80% to post-UVBF 93% PTCC) and the BL-M comparison (22% PTCC at baseline to 85% PTCC in the maintenance session).

Listeners were more confident in their responses and their reaction times were faster when they selected post-therapy versions of the word (B) as being closer to the target, however statistical analysis found no significant differences between whether listeners selected A or B. A weak negative correlation was found in the UVBF and BL-M comparisons, suggesting that when listeners were more confident in their responses, their reaction times were quicker. However, this should be interpreted with caution as Pearson’s correlations were weak for all three comparisons and this was not always significant.

3.5 Methodological Discussion

3.5.1 Sub-Study Summary

Lohmander and Olsson (2004) report that perceptual speech evaluation is the basis of speech assessment for individuals with CP. Previous reviews of the literature (Morris 1973, Dalston et al. 1988, Sell and Grunwell 2001) found methodological flaws in perceptual evaluation of the speech in individuals with CP, such as the use of different professionals as raters and a lack of inter- and intra-rater reliability for listener agreement.

These sub-studies piloted a new perceptual methodology for evaluating speech outcomes after the intervention in chapter 2, which outlines PCC and PTCC scores derived from standardised assessment and untreated target-specific wordlists. Previous literature for evaluating speech in individuals with CP suggests, for example, using two listeners with a large amount of experience in the field for phonetic transcription (Sell 2005). Whilst all listeners in these sub-studies were not expert listeners, multiple-phonetically trained listener responses were generally in-line with PTCC scores of three more experienced listeners. By conducting a perceptual evaluation using multiple-listeners along with analysis of phonetic transcription, this thesis fits with Britton et al. (2014)'s suggestion that perceptual assessment would include a robust listening procedure using multiple listeners and inter- and intra-rater reliability.

Two sub-studies, 2a and 2b, were conducted to maximise the number of comparisons and to avoid recency effects, listeners assigned to either Andrew or Craig for one of the sub-studies and the other speaker for the other sub-study. Both sub-study 2a and sub-study 2b sought to determine via a multi-listener perceptual evaluation whether listeners are able to detect any improvement in production of untreated single words presented as audio stimuli, in comparison to the target English word and with knowledge of the goal of therapeutic intervention. These were compared to PTCC scores from sub-sections 2.2.3.4 and 2.3.3.4.

3.5.2 Listener Responses

The primary aim for both sub-studies was to determine whether listeners selected a later session in time (therefore, including some comparisons of pre- versus post-therapy) as closer to the target more often than an earlier in time. Preference for the later session can therefore indicate either a rising baseline, post-therapy improvement or generalisation throughout the maintenance phase. Since the therapy design (see chapter 2) compared two interventions, and since assessments were carried out in six sessions, the perceptual method was carried out in the comparison of six different pairs of pre- and post-therapy time points- before and after therapy with VAMS, before and after therapy with UVBF, and overall improvement from baseline to maintenance; and between periods of no intervention – baseline-pre-VAM, post-VAM to pre-UVBF and post-UVBF to maintenance.

The hypothesis that listeners would select the later time-points (i.e. post-therapy) more than the earlier time-points (i.e. pre-therapy) was confirmed, with the majority of later time-points being selected as “closer to the English target” by most listeners. This was the case especially clearly in the case of Craig. In fact, listeners detected a further improvement in Craig’s speech after ultrasound even after an early improvement after the first block of VAM therapy. Clinically, this is to be interpreted with caution because he had clearly acquired the new speech sound before commencing the UVBF therapy. In other words, results here do not convincingly show that ultrasound was effective for Craig. Actually, it could have been further generalisation from the effects of therapy with VAM which showed improvements. Indeed, VBF is thought to be most useful for establishing motor programmes for new articulations (Gibbon and Wood 2010), thus probably rendering it unnecessary once Craig had learned to produce a velar articulation in the VAM block of therapy.

For Andrew, listener judgements unexpectedly indicated a decrease in understandability/accuracy/acceptability post-therapy using ultrasound (UVBF Comparison), from seven out of 10 listeners. This is contrary to previous studies reporting success with UVBF (Bacsfalvi et al. 2007; Bacsfalvi 2010; Bacsfalvi and Bernhardt 2011; Cleland et al. 2015c), highlighting the need to design larger studies which compare UVBF with competing therapies, rather than no treatment. This will help decipher what it is about UVBF that differs from competing therapies such as traditional articulation therapy/motor-based therapy, or competing tools such as VAMs. Although there is limited evidence for the use of VAMs, previous studies use only a VAM for speech reading and do not include the use of VAMs as an adjunct to motor-based therapy with the addition of explicit instruction and SLT feedback. UVBF may have an advantage over VAMs, with the additional biofeedback elements such as the child seeing their tongue moving in real-time, triggering mirror neurons for self-regulation (Cleland and Scobbie in press). The other advantage of UVBF may be the delayed feedback of being able to watch recordings of attempts at a target.

For both children, phonetic transcription showed an increase in percentage of targeted consonants correct from initial baseline to maintenance, three months after therapy ceased. For Andrew, this improvement was modest, rising from 5% PTCC at baseline to only 21% PTCC at maintenance. In sub-study 2b for the post-UVBF-maintenance comparison, 55% of listeners thought that maintenance sounded better. This would be expected due to the increase in PTCC from 5% in post-UVBF to 21% in maintenance. However, this is unlikely to represent a clinically significant improvement in Andrew's production of /n/, in line with Preston et al. (2014)'s benchmark of 20% for clinical significance, suggesting that neither therapy was particularly effective.

In contrast, Craig improved from 22% PTCC at baseline to 90% PTCC at his maintenance recording, more strongly suggesting he had successfully integrated velars into untreated words. 60% of listeners selected B more than A in the post-UVBF-maintenance comparison, suggesting further generalisation 3-months after therapy ceased, however PTCC scores remained stable at 93% in the post-UVBF session and 90% in the maintenance session. In fact, two out of three of the transcribers in inter-rater measures showed stability between post-UVBF and maintenance with one of the blind transcribers showing a decrease in PTCC scores in the maintenance session. This might suggest that this method of perceptual evaluation is useful for detecting subtle improvements of speech. Due to the mismatch in results from phonetic transcriptions and perceptual evaluation in this instance, it is then useful to investigate the articulatory data (see chapter 4) to confirm whether there are in fact any further changes in lingual patterns in the maintenance session.

Despite previous literature suggesting that transcriptions from single transcribers are unreliable and multi-listener judgements being preferable (Kuehn and Moller 2000; Lohmander and Olsson 2004; Britton et al. 2014) results of this novel methodology closely corroborate aspects of the phonetic transcription. When listeners select B (later/post-therapy) as closer to the target, the PTCC score also increase and likewise, if PTCC scores decrease, as in Andrew's UVBF comparison, PTCC scores also decrease.

However, with the differing methodologies of the perceptual evaluation and the phonetic transcriptions it was not possible to correlate results statistically. Previous literature suggests that point-by-point reliability for broad phonetic transcription is often in the 90-95% range and for narrow transcription is often around 80% (Shriberg and Lof 1991; Shriberg, et al. 1997). Preston, et al. (2011) point out that it is more difficult to achieve agreement on disordered speech, with complex speech disorders such as those found in cleft palate often being associated with low inter-rater agreement (Shriberg and Lof 1991). Gooch, et al. (2001) found an average of 40% agreement across listeners (range 19%-71%) when comparing listener judgements against transcriptions of compensatory articulations. Based on the range of reliability proposed by Shriberg and Lof (1991) and Shriberg et al. (1997), results would suggest that the average of 74% accuracy across both speakers is not reliable, highlighting the need for multiple listener perceptual evaluations such as this.

In terms of raw measures, listeners were more confident and quicker to respond when selecting B as closer to the target for both speakers; however, results were not significant with the exception of the UVBF comparison for Andrew where listeners were significantly more confident when they selected A (pre-therapy). Although there was a trend between quicker reaction times and higher confidence, correlations were weak for both speakers in all comparisons and mostly non-significant.

In terms of agreement between listeners, statistical analysis showed varied levels of agreement, with listener judgements matching the PTCC scores derived from phonetic transcriptions in subsections 2.2.3.4 and 2.3.3.4. Listener judgements were more reliable for Craig than Andrew, which is probably the result of chance levels (i.e. guessing) when tokens from different time-points were indistinguishable. In this sense the perceptual evaluation is quite different from a phonetic transcription which is not designed to detect improvement without further analysis (for example calculating PCC). Moreover, with double articulations suspected through narrow phonetic transcriptions there was ambiguity in productions. In the case of double articulations, a VAS may have been useful to detect whether listeners heard the token in question as being more velar-like or more alveolar-like. However, this would assume that listeners have prior knowledge of place of articulation for /t/ and /k/. The protocol presented in this sub-study does not require any articulatory knowledge and the comparative design measures only whether one token is essentially better than another. Using phonetically trained listeners has its advantages, with previous literature suggesting specialist SLTs are more reliable than non-specialist SLTs (Keuning et al. 1999). While the listeners in this sub-study were not specialist SLTs in the field of CLP, they were still phonetically trained. However, the methodology presented in this sub-study also allows for use with naïve listeners, with no phonetic knowledge necessary, which may further add to the validity of acceptability measures and provide real-life significance to clinical speech assessments.

Whilst poor listener/transcriber agreement may be evident through statistical analysis, it should be noted that the kappa is a conservative statistical measure. It assumes a high level of agreement obtained by chance when judgements are not evenly distributed (Cordes 1994; Brunnegard and Lohmander 2007). In other words, where there is a high agreement in listener responses, due to a ceiling effect, there is a higher chance of a low kappa score.

3.5.3 Feasibility

The perceptual evaluation in this sub-study provides a relatively quick and easy method of testing pre- and post-therapy speech with multiple-listeners for research purposes. From a practical perspective, the MFC document in PRAAT version 5.3.57 (Boersma and Weenink 2013), is easily modified by copying and pasting audio file names into the document. This process is quick and easy for research purposes and takes no longer than 60 minutes to complete. Since conducting the study, three small projects have replicated the methodology with ease (Alexander 2015; Thompson

2015; Young 2015). These studies used audio data of children with primary SSD from a different project using UVBF (ULTRAX, 2011-2014) to evaluate speech outcomes pre- and post-therapy. Whilst it may be quick and easy for research purposes, in clinical practice this may not be a process that can be easily or quickly designed or administered in a busy clinic.

Although the current study used phonetically trained listeners who have experience in listening to disordered speech, the methodology is designed so that lay listeners can also be used. Listeners were asked to select which version of a word was "closer to the English target" based on their own representation of a word from phonological and phonetic intuitions, which does not require phonetic skill. Future studies using this perceptual evaluation protocol should compare ratings by expert and lay listeners. Future studies should also compare this comparative method to other methods, such as VAS. Magnitude measures, such as VAS and direct magnitude estimation (DME), involve listeners assigning numbers to stimuli in proportion to their magnitude (Yiu and Ng 2004). VAS may be a useful method in the case of velar-alveolar double articulations, to identify whether listeners perceive for example, how velar-like or alveolar-like a token may be. This may circumvent some of the issues with ambiguity in phonetic transcriptions and supplement listeners' perceptions of which version is "better" pre- and post-therapy. It would be expected that if VAS would be compared to the comparative method presented in the current sub-study, that listeners would perceive a target as "more velar-like" on the scale when they select a token as being "closer to the English target".

Furthermore, although no obvious differences were found in the current study between the PTCC scores and the perceptual evaluation, in a master's thesis using the same methodology to evaluate pre, during and post-therapy changes in a child with Childhood Apraxia of Speech (CAS), listeners identified subtle improvements between mid-therapy recordings and post-therapy recordings, both of which were rated as 100% on target by a transcriber (Young 2015). All in all, these results suggest that this perceptual evaluation method might be useful for detecting subtle improvements in acceptability, without the need for time-consuming narrow transcription. This methodology is suitable not only for perceptually evaluating data from speakers with CP but also those with a range of primary or secondary SSDs.

3.5.4 Methodological Limitations

In this study the perceptual evaluation was piloted only on phonetically trained listeners. This had the advantage that it was straightforward to explain to listeners which sound they should focus on in the audio stimuli. However, previous literature states the benefits of using lay listeners in perceptual evaluation of the speech of individuals with CP. Although listeners had little to no experience of working with clients with CP, all had completed phonetic training. Gooch et al. (2001) report the difficulty associated with fine phonetic transcription of compensatory articulations in speakers with CP. With two groups of transcribers (group 1 – SLTs experienced in evaluating speakers with CLP; and group 2 – SLTs not experienced in evaluating speakers with CLP), they found that only 40% of transcriptions, on average, agreed with expert's transcriptions. Experienced listeners, on a whole, agreed more than inexperienced clinicians; however not as well as expected. Keuning et al. (1999) also found that experience in listening and analysing the speech of individuals with CP did not ensure higher inter-rater reliability.

Despite this, Sell (2005) suggests that the level of experience of the listener should be considered as an index of reliability when designing perceptual evaluations. It would, therefore, be beneficial to further test this methodology comparing expert listeners (i.e. those with extensive experience working with CP) to lay listeners. Using lay-listeners would also have the advantage that is might be possible to employ this methodology using remote listeners via the internet or even using Crowdsourcing. McAllistier Byun et al. (2015) found that it was possible to use lay-listeners to rate the speech of children with mild articulatory difficulties (/r/ misarticulations) using the Crowdsourcing platform "Amazon Mechanical Turk" quickly and easily. Using our method, it would be possible to do the same for more severe SSDs, especially where speech is less intelligible.

Another methodological limitation is that the loudness levels/recording quality was not controlled during assessment sessions and was therefore different for each session. As this was not altered for the perceptual evaluation, this could have led to listener bias toward one session if the recording was of clearer quality. Future studies should control for the volume and clarity of the recordings where possible.

3.6 Summary of Perceptual Evaluation

In summary, the perceptual evaluation shows promise as a method of evaluating speech outcomes from any speech therapy such as ultrasound and visual articulatory models. In this case, the evaluation showed substantially improved speech in one speaker (Craig) and little gain for the other (Andrew). These results mirror the PTCC scores in sub-sections 2.2.3.4 and 2.3.3.4 despite weak listener agreement. Correlations between reaction time and confidence were weak and mostly non-significant.

Using a similar methodology for perceptual evaluation of speech in a case of CAS, Young (2015) detected subtle differences mid-therapy, similar to those detected for Craig between post-UVBF and maintenance. While the perceptual evaluation highlights some subtle differences in speech outcomes and is able to tell us which of two versions sounds closer to a target, it does not tell us what these changes or differences are or when a token is not close to a target, what the error is. The phonetic transcriptions provide detail on these changes and errors. However, issues with perceptual assessment have been highlighted (e.g. Sell 2005), with the possibility of covert errors, particularly in the complex speech of individuals with CP. Instrumental articulatory analysis, for example using EPG (Gibbon 2004) or UTI (Gibbon and Wolters 2005; Bressmann et al. 2011) may supplement the perceptual data by providing additional information on covert errors, such as double articulations, palatalisation, or retraction. The following chapter will present articulatory data from ultrasound tongue images in order to address some of the difficulties highlighted in the perceptual analysis present in the current and previous chapters.

4 Articulatory Analysis

This articulatory analysis chapter contains a method section followed by a results section presenting results from Andrew then Craig. These results provide a specific articulatory analysis (drawing on synchronised ultrasound and audio recordings). In fact, the audio part of these recordings is the same data previously reported from the perspective of perceptual evaluation and transcription-based analysis in chapter 2. Articulatory data provides additional information, which supplements the phonetic transcriptions. The general data collection method with materials is reported in section 2.1. The following method section, therefore, presents the method for analysing ultrasound data.

4.1 Articulatory Analysis Method

4.1.1 The Purpose of Articulatory Analysis

Firstly, it is important to analyse data instrumentally using ultrasound, because ultrasound can not only be used for visual feedback in therapy but also diagnostically to determine therapy targets. Secondly, it was used to compare post-therapy outcomes to pre-therapy tokens, to determine whether there was any difference post-therapy and if so, how big are the differences and do they reach significance levels. Further, from a methodological perspective, it is important to consider whether ultrasound is a practical tool for analysing longitudinal data, which can be concluded from the measurements made in pre- post-therapy comparisons. Specific research questions for Andrew and Craig's data are outlined below in sub-sections 4.1.2.2 and 4.1.2.3.

4.1.2 Ultrasound Analysis and Outcome Measures

A qualitative analysis of ultrasound data was carried out for each of the six assessment sessions for both children. Additionally, quantitative analyses were implemented to identify tongue surface length and width and statistical differences within minimal pairs for Andrew and between alveolar and velar tongue shapes for both Andrew and Craig to investigate their therapy targets (Andrew: /n/, Craig: velars). One of the potential challenges for quantifying and interpreting articulatory data is the lack of existing information about how the relevant consonants are articulated by adults and typically-developing children. It is possible to draw on pre-existing data from a number of projects at Queen Margaret University in order to provide some context and provide a comparison of tongue shapes and typical values for factors such as tongue length for alveolar and velar plosives. Tongue length refers to the length of visible tongue between the mandible and hyoid shadows. In typically developing (TD) children, it is likely that some of the tongue-tip data may be missing due to a large mandible shadow. Previous studies have shown that the average visible tongue length for /a/ in children aged 7;7 is around 6cm (Zharkova 2012). By measuring the length of visible tongue in Andrew and Craig's data, this allows for a comparison to TD peers, to determine whether the quality of the image (the amount of tongue visible on the screen between the two shadows) is of better or poorer quality than TD peers. This will help to consider whether children with CP and accompanying syndromes (e.g. Pierre Robin Sequence where a large mandible shadow would be expected due to small jaw) are suitable candidates for ultrasound tongue imaging. However, as there are only two participants here, results cannot be generalised to all children with CP.

4.1.2.1 Image Quality: A Comparison to Typical Data

Firstly, a qualitative comparison of raw images will be presented, which will provide an overview of the quality of the images for both Andrew and Craig compared to age-matched, typically developing peers. Secondly, quantitative measures from typically developing children for the length, area and width of the tongue will be provided.

Typically, ultrasound data is recorded with tongue tip to the right. However, as this thesis compared ultrasound to a VAM which had tongue tip to the left (see Figure 8 in section 1.5.3 of the literature review), it was appropriate to be consistent with orientation for participants. Therefore, for the purpose of this thesis, ultrasound data was recorded with tongue tip to the left of the image. Figure 47 shows an example of data from a typically developing child's data for an alveolar [n]. Labels of the different regions of the tongue are provided. It is important when recording to have the mandible and hyoid shadow as equal as possible on the image, ensuring that the maximum portion of the tongue is visible.

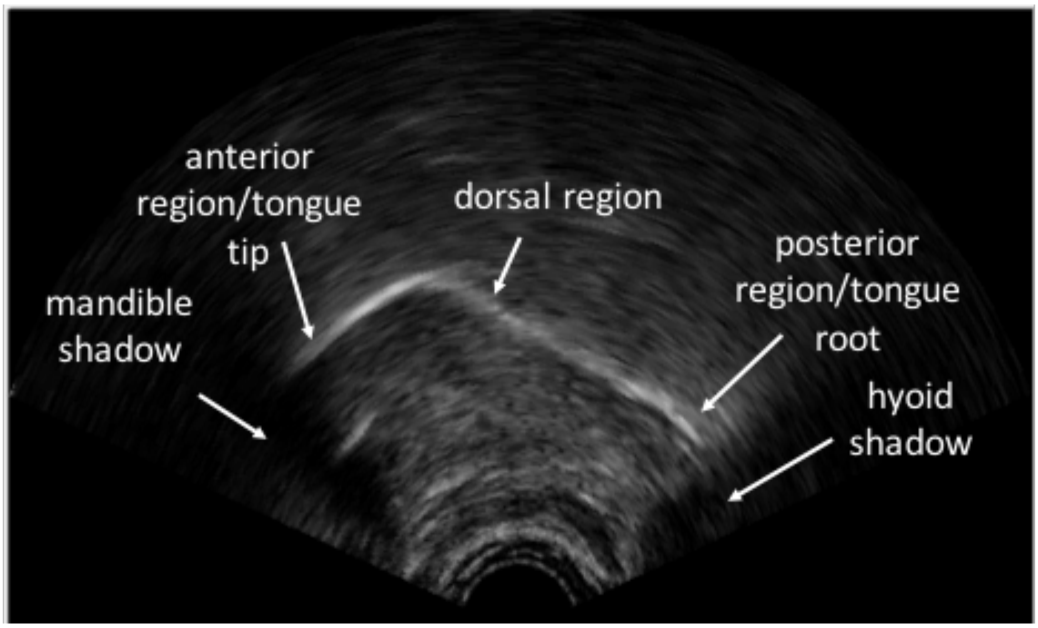


Figure 47 Orientation to ultrasound image (example of alveolar /n/ from ULTRAX Project)

Age-matched typically developing peers were selected for Andrew and Craig from data of 30 children derived from the ULTRAX project (2011-2014). Qualitative analysis of image quality was carried out to help evaluate the potential for quantitative analysis and to help with interpretation. The quantitative measures of typically developing children reported here were analysed collaboratively (tongue length, by the author and Scobbie) or derived from annotation and edge tracking (width and area, Scobbie and Cleland 2017).

Eight children (three male; five female), aged between 6;8 and 7;11, were selected as an age match for Craig and seven children (three male; four female), aged between 9;4 and 10;8, were selected for Andrew. From these age-matches, a participant with good quality data (i.e. a clear tongue surface and equal mandible and hyoid shadows) and a poor-quality image (i.e. missing data or artefacts) was chosen to compare to the raw images of Andrew and Craig's ultrasound data for their target consonant (Andrew, /n/; Craig, velars).

Figure 48 shows a comparison of Craig's velar target to that of two typically developing males, aged 6;8 and 7;11. It is clear that the quality of the images for Craig are of poorer quality to that of the male aged 6;8, with a closer representation to that of the male aged 7;11 with poor image quality. In Craig's data, there is a large mandible shadow and the mandible and hyoid shadows are not equal. This results in a small portion of the tongue being visible. This is likely due to Craig's small chin. See section 4.2.2 for qualitative and quantitative analysis of Craig's data.

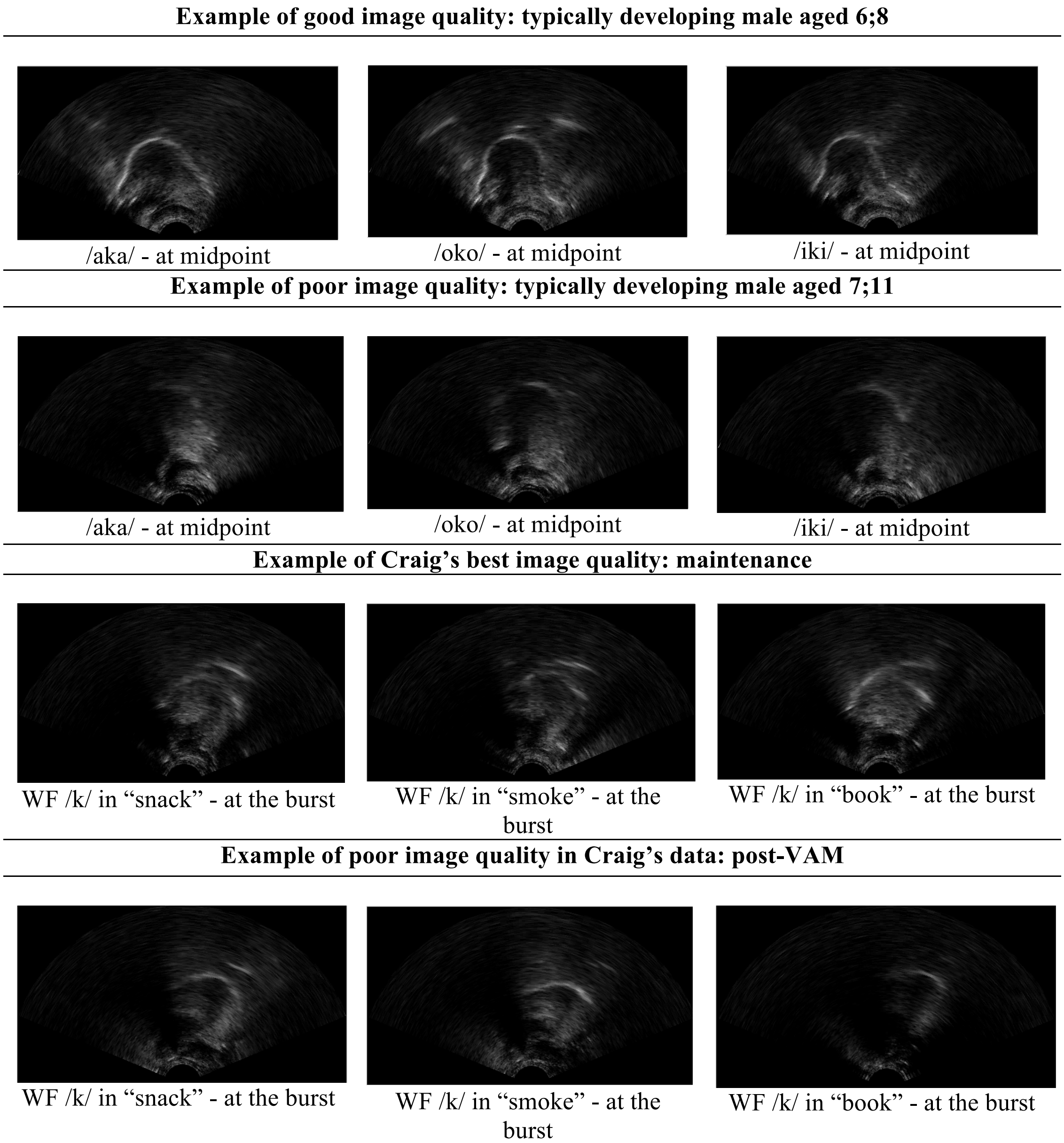


Figure 48 Comparison of Craig's raw image quality of target /k/ to that of age-matched, typically developing peers (Anterior to left)

It is also important to consider the quality of the dynamics. Figure 49 shows a series of raw ultrasound images taken from a single recording in Craig's baseline session. Within the recording, the hyoid shadow moves and artefacts appear on the screen, indicating headset movement and the possibility that the probe is not in a midsagittal position. As Craig's head was small, this allowed the headset, primarily designed for adults, to move more easily than is desirable.

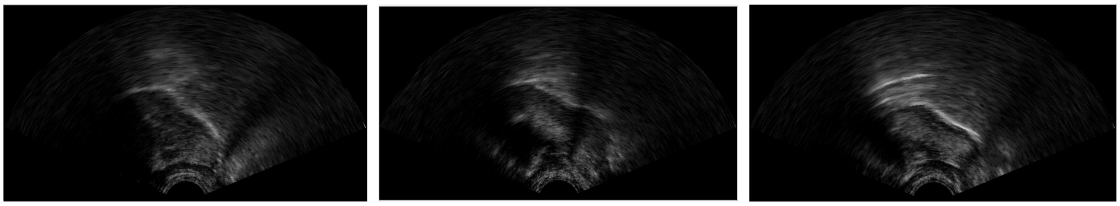
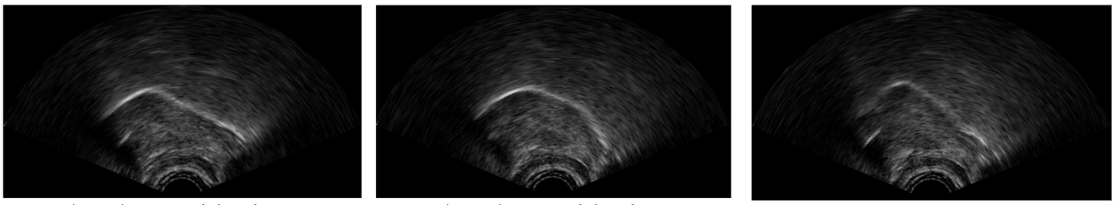


Figure 49 Series of Images from one recording for Craig at baseline showing headset movement

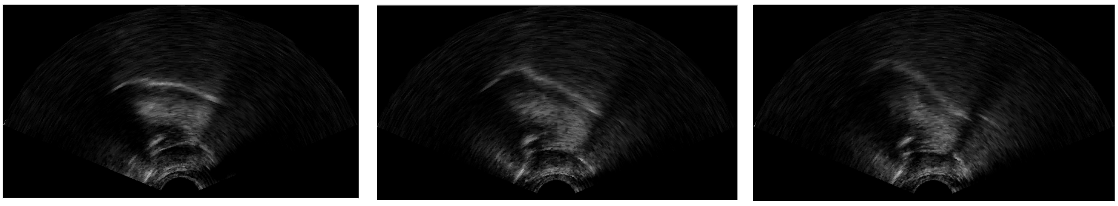
Turning now to Andrew, Figure 50 below shows a comparison of Andrew's target /n/ to that of two typically developing females, aged 9;4 and 10;5. It should be noted that Andrew's productions of /n/ in the images are incorrect and therefore the tongue shape is not being compared here, only the quality of the image. Although not as clear as in Craig's data, there is a larger mandible and hyoid shadow in Andrew's data than in the image of the female aged 9;4. The good image quality for Andrew in the post-UVBF session is not dissimilar to the poor image quality of the female aged 10;5. However, the image for Andrew in the pre-UVBF session is of particularly poor quality, with a small portion of the tongue visible and a fading image with elevation of the tongue, so this session needs to be approached with caution.

Example of good image quality: typically developing female aged 9;4



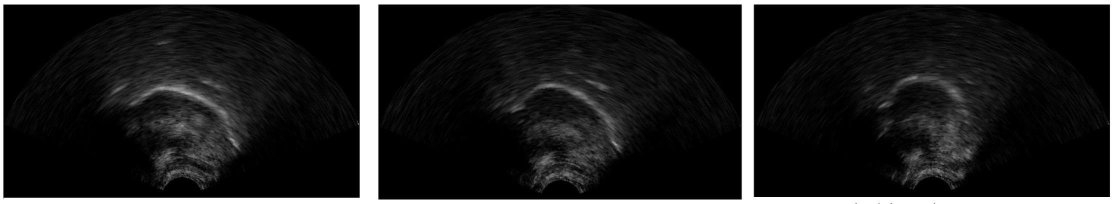
/ana/ - at midpoint | /ono/ - at midpoint | /ini/ - at midpoint

Example of poor image quality: typically developing female aged 10;5



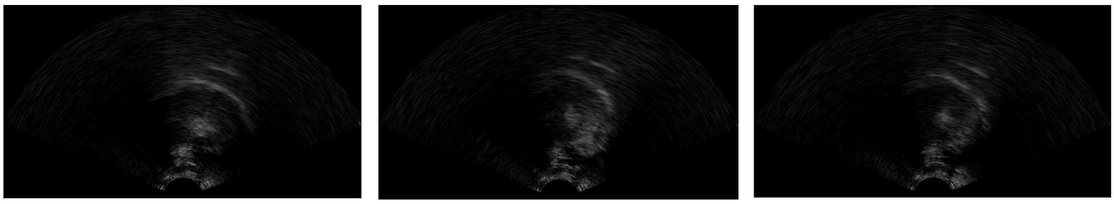
/ana/ - at midpoint | /ono/ - at midpoint | /ini/ - at midpoint

Example of Andrew's best image quality: post-UVBF



WI /n/ in "nap" – at midpoint | WI /n/ in "know" – at midpoint | WI /n/ in "knee" – at midpoint

Example of poor image quality in Andrew's data: pre-UVBF



WI /n/ in "nap" – at midpoint | WI /n/ in "know" – at midpoint | WI /n/ in "knee" – at midpoint

Figure 50 Comparison of Andrew's raw image quality of target /n/ to that of age-matched, typically developing peers

With variability evident in Andrew and Craig's data across sessions, this is likely to be due to the probe being in a slightly different orientation within each session. The headset was in fact particularly difficult to fit correctly on Andrew compared to other children, due to his hemifacial macrosomia and unilateral microsomia. Similar to Craig, Andrew also had a small space under his chin. Thus, facial symmetry and jaw size should be considered when selecting suitable participants for ultrasound based intervention. The poor image quality was of particular difficulty because for both participants, particularly Andrew, the therapy target included alveolar placement. The large mandible shadow means that tongue tip data, crucial for alveolar consonants, is missing. Therefore, it was difficult to provide instruction and to use the ultrasound as biofeedback for alveolar consonants.

Whilst it is helpful to look at image quality in terms of the raw data in a qualitative fashion, quantitative measures are crucial in the analysis of ultrasound data, particularly that of any clinical population, to identify any covert errors or covert contrasts in data, such as those found in Cleland et al. (2017b). Scobbie and Cleland (2017) provide norms with which to compare incorrect /t/ and /k/ productions. They used quantitative analysis to measure the dorsal constriction in the /t/-/k/ contrast. Data from 30 typical children (ULTRAX 2011-2014) of single tokens of /k/ and /t/ in three VCV contexts (between symmetrical /a/, /i/ and /o/) was analysed to measure what they call the KT crescent, both the width (i.e. the maximum radial difference between /t/ and /k/, and the area (the sum of annular sectors, each centred on a single fan line). The KT crescent is intended to quantitatively measure velar and alveolar tokens from the same speaker, and measures the area in which the two tongue-curves cross over. For example, in Figure 51, the image on the left shows two tongue curves (/t/ and /k/) with two overlaps. The KT crescent on the right, is the area between the two points of crossover. From this, the width and area of the crescent can be calculated.

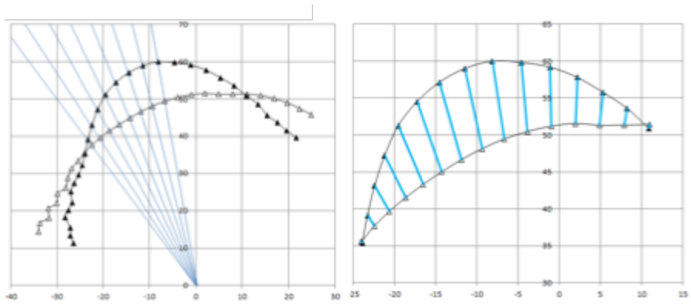


Figure 51 Example of a KT Crescent – tongue tip on the right. Used with permission from Scobbie and Cleland (2017)

They found that the width and the area of the contrast between /t/ and /k/ are highly correlated, meaning that either measure can be used as a comparison for the typical velar vs. alveolar difference. Table 55 shows the minimum and maximum area and width for the three vowel environments, with the mean width and area denoted by bold-case (used with permission from the authors). Scobbie and Cleland suggest that the width and area stay roughly the same independent of the ages of the child.

	/a/	/i/	/o/
Area (mm^2)	**294**	**164**	**272**
min	159	43	52
max	450	405	554
Linear(mm)	**11.9**	**7.5**	**12.1**
min	7.3	3.3	5.9
max	18.0	16.0	22.0

Table 55 Mean, maximum and minimum width and length for the /t/-k/ contrast (Scobbie and Cleland 2017)

The stability of these KT crescent measures is not likely to be simply due to lack of data, given other developmental trends that they were able to measure from this data set. They found that tongue length increases each year, by 2.6mm for /k/ and by 3.0mm for /t/, and also observed that the distance of the tongue surface in the middle of a /k/ increases by approximately 0.8mm per year.

As both Craig and Andrew were making errors with alveolar and velar stops, the ULTRAX data, and the measurements from Scobbie and Cleland (2017), made it possible to calculate an expected tongue length for age-matched peers for both Andrew and Craig, for /t/ and /k/ (see below), to compare the quality of the data of children with submucuous cleft with the data quality of TD peers. It is to be expected, based on previous literature (e.g. Gibbon et al. 2007), that nasal and oral stops should be the same, in terms of place of articulation and tongue shape. Therefore, the norms for oral stops will also be applied to nasal stops in Andrew's case. Another value of using the measures proposed by Scobbie and Cleland (2017) is that it allows for longitudinal comparisons, even if the headset is not in the same orientation within each session.

On Scobbie and Cleland's linear model, the observed tongue length formula for /t/ is increasing by an average of 3mm per year, therefore the length of /t/ (mm) = 3 x age (years) + 35 (offset). For /k/, it increases by 2.6mm per year, therefore the length of /k/ (mm) = 2.6 x age (years) + 39 (offset). Table 56 provides the tongue lengths expected for ages six, seven, nine and 10 years, as these are the ages of Craig and Andrew pre- and post-therapy, allowing for age-matched comparisons.

Age (years)	Tongue Length /t/ (mm)	Tongue Length /k/ (mm)
6	53	55
7	56	57
9	62	62
10	65	66

Table 56 Tongue length norms for age-matched peers

Zharkova (2012) reports that the visible tongue length observed for the vowel /a/ is around 6cm at age 7;7, which is not too dissimilar to these figures.

Tongue length data is based on the surface of the tongue that is visible. As seen in Figure 48 and Figure 50 above, even in typical data, some of the tongue-tip image can be missing due to the large mandible shadow. It may be expected in some speakers with CP, particularly those associated with conditions with a small jaw, such as Pierre Robin Sequence, that there will be a large mandible shadow on the data and therefore, tongue tip data may be missing. Though not all of the tongue will be visible, the values that are derived from ultrasound may or may not be reliably representative of the actual full length of the tongue, nor is it clear what a reasonable estimate of tongue length is, from the ultrasound data. Both Craig and Andrew had a small jaw, therefore it may be expected that the tongue length values will be shorter than those in the data of typical children.

The following sections will provide a method for qualitative and quantitative analysis of articulatory data for both Andrew and Craig. The results that follow will be compared to the typical norms from Scobbie and Cleland (2017) throughout.

4.1.2.2 Andrew

The phonetic transcriptions presented in section 2.2.3 showed that Andrew was retracting his alveolar nasal /n/ to a velar placement, with some ambiguity suggesting possible double articulations. Both transcriptions and perceptual evaluation showed an overall improvement in speech outcomes. Ultrasound analysis was used to determine:

1. Whether /n/ showed a clearly retracted and typically velar shape.

2. Whether there was a merger between /n/ and /ŋ/ or whether a covert contrast was present indicating a phonetic-level impairment. If a covert contrast seems evident, the quantitative differences in tongue shape, length and width between /n/ and /ŋ/ are used to provide statistical justification of this.
3. Whether there was a change in tongue shape post-therapy, in each of the three comparisons (VAM, UVBF and BL-M), therefore indicating improvement post-therapy.

Firstly, using AAA v2.16 software (Articulate Instruments 2015), single words from the untreated wordlist were annotated. Whole /n/ segments were annotated using spectral and waveform characteristics and then the midpoint of /n/ were identified using an automatic function. The corresponding (within half a frame rate) midpoint ultrasound frames were then selected and splines indicating the tongue surface were fitted to the images using the semi-automatic edge-detection function in AAA software. Splines from /n/ in different word positions (WI, WM and WF) and in a range of vowel environments were then averaged and compared within session in the AAA workspace. Figure 52 and Figure 53 give examples of multiple tokens of /n/ and an averaged spline with standard deviations denoted by dashed line in Figure 53, taken from Andrew's baseline productions of /n/.

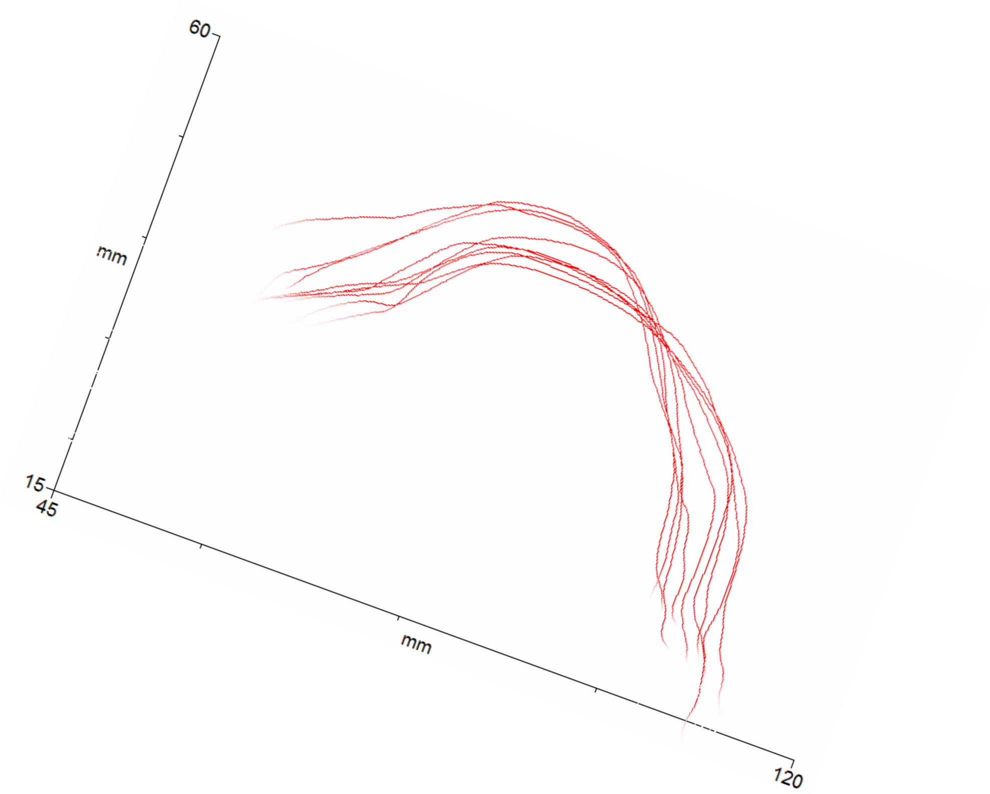


Figure 52 Example of multiple tokens of /n/ taken from baseline (rotated at 20°)

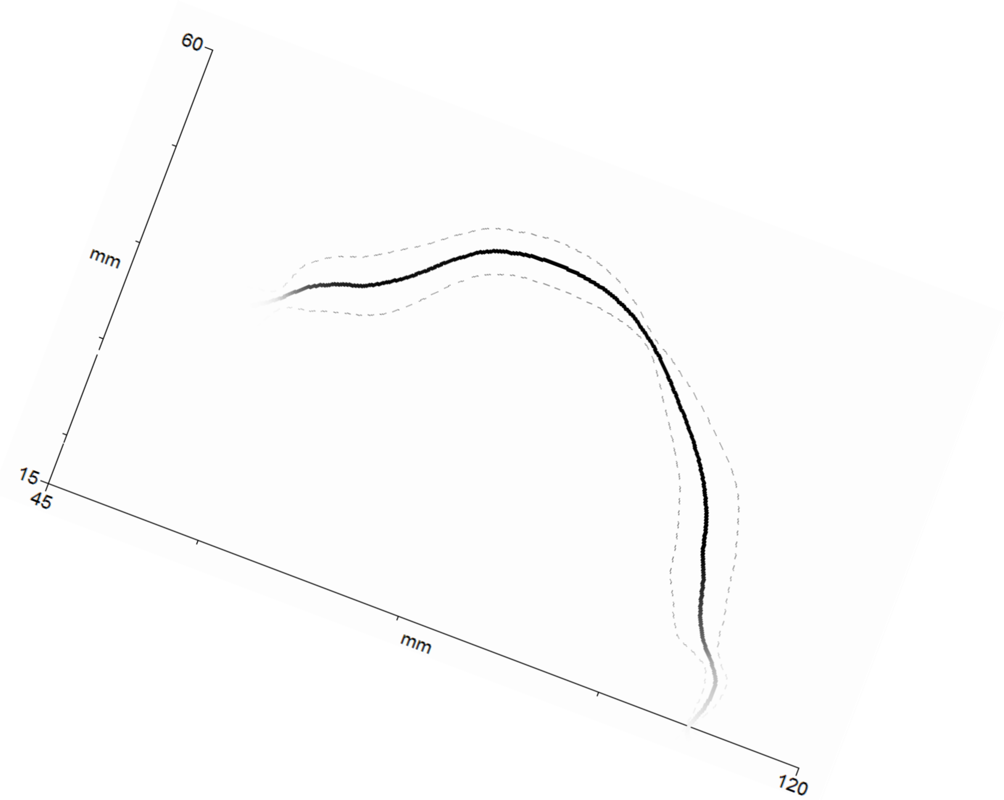


Figure 53 Example of averaged /n/ from the multiple splines presented in Figure 52 (rotated at 20°)

The figures above show how data from one session can be averaged more easily for statistical testing. Longitudinal data on the other hand presents a new problem. It can be difficult to align across sessions due to the probe being placed in a different orientation in each session. To circumvent some of the issues around aligning longitudinal data, there are various methods that can be used. Firstly, a palate trace can be obtained by asking participants to swallow, with the reflected image being that of the tongue pressing against the hard palate. Palate traces from each session can then be aligned in the AAA workspace, in turn re-aligning all tongue images into the new orientation (Cleland et al. 2015c). Another method for aligning data is to use a bite plane (Scobbie et al. 2011). Figure 54 shows the rotation and translation of palate traces and tongue splines in AAA workspace, using the move and rotate functions in the bottom right hand corner. Figure 55 shows palate traces before and after rotation and translation. Data from Figure 54 and Figure 55 is taken from the ULTRAX Project (2011-2014) in which the author was involved in analysing the data for a child who was velar fronting. Note that in Figure 54 and Figure 55 the tongue tip is to the right.

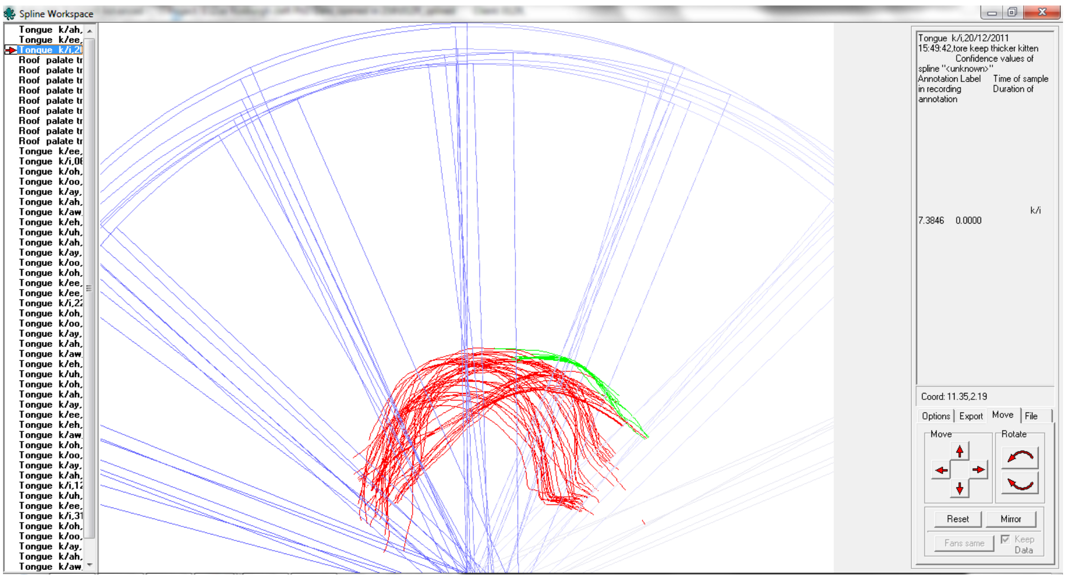


Figure 54 Rotation and Translation of Palates (green) and tongues (red) in a child with SSD: velar fronting. Tongue tip on the right.

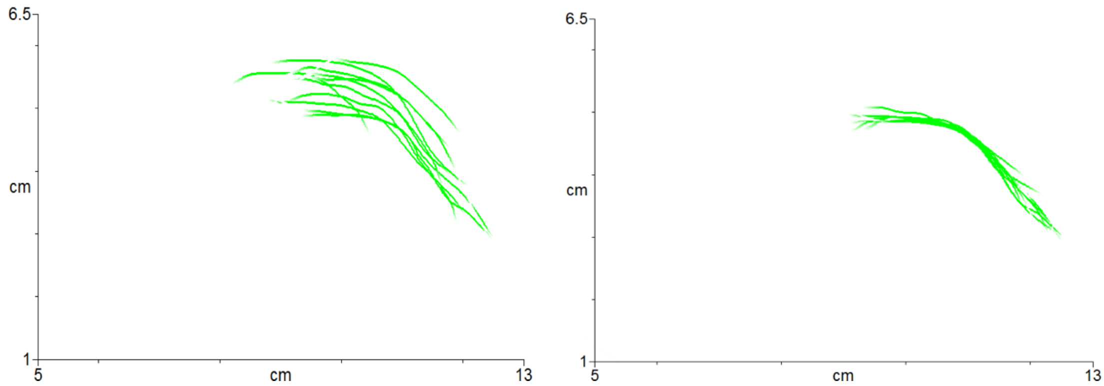


Figure 55 Alignment of Palate Traces in a child with velar fronting (left: unaligned; right: aligned). Tongue tip on the right.

In the current study, data was not compared across assessment sessions as there was no palate trace visible to align the data across sessions and each session the probe was in a different orientation. A bite plane was not used as there were no available bite planes suitable for children's mouths at the stage of data collection.

Another means to deal with probe movement between sessions is to generate a quantitative value within each session, for example using a paired t-test to compare two tongue curves (e.g. of the target /n/ and the minimal pair token /ŋ/), and then compare those values longitudinally. The /n/, /ŋ/ minimal pairs from the additional alveolar wordlist provided tokens of matched /n/ and /ŋ/ segments.

Similar to previous studies (Cleland et al. 2015c; Cleland et al. 2017b), in order to quantify differences between /n/ and /ŋ/, the built-in statistical difference function in AAA was used. Within the workspace on AAA, all tokens of a segment can be averaged by selecting all tokens and pressing the "mean & s.d." button. By averaging two segments (e.g. /n/ and /ŋ/), the two average tongue curves can then be compared statistically using the built-in statistical difference function by pressing "Diff". Using this function, significance is tested along 42 fan-lines radially from the probe to the tongue surface. The mean difference for a given fan-line is treated as significant at $p < .05$, but it is important to understand that the presence of one significant t-test of even a high level (e.g. $p<.001$) within a range of non-significant fan-lines should not be interpreted as more important than a span of several contiguous fan-lines just passing the threshold of $p = .05$. In their "zone of significance" Cleland et al. (2017b) use six adjacent fan-lines as the spatial threshold to report a significant difference between two sets of tongue shapes. They also discuss the incorporation of “cross-overs", where two mean splines meet, to maximize the zone of significance. The current study uses the criteria of five adjacent fan lines with significance, with a maximum of five fan-lines in the cross-over region for the zone of significance. Five adjacent fan lines were chosen as there was only one dataset with six or more fan-lines in the cross-over region as in Cleland et al. (2017b). All other datasets had maximum five adjacent fan-lines in the cross-over region.

As well as reporting whether there was such a global significant difference between two tongue shapes, it is important to determine how big the difference is. Even if there is no zone of significance, it is possible to quantify the relative similarity of two tongue splines that may have insufficient numbers of significantly different radii, or none. This has the potential to allow tracking of non-significant changes in how similar /n/ and /ŋ/ are. When a zone of significance appears e.g. after treatment, the average difference between the splines would have increased too. The reverse is not true, but plotting the quantitative size of the difference between two sets of splines independent of their significance may be a useful subtle measure of change. Thus, it is important to quantify raw similarity/difference and to quantify the size of significant differences, if there are any.

Two quantitative measures of the similarities/differences of two tongue shapes are used. First, the average and maximum distance (width) between the splines is reported (measuring along the radii). This differs from Scobbie and Cleland (2017) in that these values are not dependent on a crescent between two tongue curves; instead they are based on comparable mean splines of two tongue curves. Therefore, the mean width from data presented in the current study may not be directly comparable to Scobbie and Cleland (2017), as the mean or maximum width may not occur in a crescent-shaped difference, and so only a small difference in width is expected. Second, the length of tongue surface is estimated. In the case where there is a zone of significance, its width and length are reported, along with the proportion of the whole tongue length involved. Where there isn’t such a zone, the average width is reported along the whole visible tongue length (with splines trimmed to be comparable), and this length reported. The average width of two non-distinct splines can be expected to be smaller than the average width of two significantly different splines in the zone of difference. From the t-test, the average difference between /n/ and /ŋ/ was calculated. From the average spline data and t-tests, the length of the tongue was modelled as a sum of a series of arcs, the length of each of which is based on the radius and the formula $2\pi r$. Each radius represents 0.9% of a circle (134.8°/41), that is 134 divided by the number of equal spaces between the 42 fan lines.

In order to address the aim about longitudinal comparability, two sounds already in Andrew's phonetic inventory (/t/ and /k/), where the tongue shape properties are already known, were compared. In this way, it is possible to evaluate them. Data from typically developing children (Scobbie and Cleland 2017) identifies that there are large coarticulatory effects for /t/ and /k/. It is, therefore, important to compare alveolar and velar plosives in different vowel environments (e.g. /a/ /i/ /o/), which can then be compared to Andrew's production of /n/ to identify the similarities and differences in /n/ with /t/ and /k/. Gibbon et al. (2007) suggests that in typical adults, there are no differences in placement for alveolar oral and nasal stops. Thus, if Andrew's productions are correct, it would be expected that there would be no statistical differences between /n/ and /t/. If his productions are in velar placement, as phonetically transcribed, it would be expected that the tongue shapes would be more similar to those of /k/.

Single tokens of WI /t/, /k/ and /n/ in 'knee, know, nap', 'tea, toe, tap' and 'key, co, cap' were annotated (/k/ and /t/ at the burst and /n/ at midpoint, similar to Cleland et al. (2015c) and Cleland et al. (2017b)) and corresponding key frames were splined. Gibbon and Wolters (2005) used the midpoint for vowels; however they used the point of closure for plosives. In the current study, the burst and midpoint were chosen as these points are reliable to find. Tokens of /t/, /k/ and /n/ were exported for comparison across vowel environments /i/, /o/ and /a/ within session. Therefore, three comparisons of the single tokens of /t/, /k/ and /n/ were created within each session: 'knee, tea, key', 'know, toe, co' and 'nap, tap, cap'.

Further, all 50 items from the DEAP were annotated for relevant tokens. In addition to the tokens of /n/, /ŋ/, /t/ and /k/, it was decided to also annotate and spline /s/ and

/ʃ/ to enable further means to compare /n/ to correctly produced alveolar targets and also to compare /ŋ/ to /ʃ/, since /ŋ/ was fronted towards the palatal region in the minimal pairs. All tokens of /t/ (N=12) and /k/ (N=12) were annotated at the burst and all tokens of /s/ (N=14), /ʃ/ (N=4), /n/ (N=8) and /ŋ/ (N=4) were annotated at midpoint, with corresponding key frames splined. Splines were averaged for /t/, /k/, /n/, /ŋ/, /s/ and /ʃ/ and compared within each session. Quantitative measures will show whether there is a statistical difference between /t/ and /n/, and between /k/ and /n/. If there is statistical difference between /t/ and /n/, but not /k/ and /n/, this will indicate that there is a velar placement for Andrew's incorrect production of his alveolar nasal stop. It would be expected, as mentioned, that there should be no statistical difference between /t/ and /n/, if Andrew's productions are correct.

To quantify these differences between /t/ and /n/, and /k/ and /n/, the built-in statistical difference function in AAA was used. Similar to the minimal pairs, the average differences between /t/ and /n/, and /k/ and /n were calculated. From the average spline data and t-tests, the length of the tongue was modelled to identify the average tongue length and the portion of tongue that was statistically different. The maximum and mean width was also calculated (see sub-section 4.2.1.4).

4.1.2.3 Craig

The phonetic transcriptions presented in section 2.3.3 showed that Craig was either retracting his velar plosives to glottal placement or fronting them to an alveolar placement in the case of the voiced plosives. Again, there was some ambiguity in the transcription, suggesting possible double articulations. Both transcriptions and perceptual evaluation showed improvement in speech outcomes at the end of both blocks of therapy and overall.

Ultrasound analysis was used to determine:

1. Whether there are any covert errors in productions of velars, i.e. when a glottal stop is identified through phonetic transcription, is it possible to check

with ultrasound if there are (appropriate or inappropriate) lingual movements?

2. Whether there are any quantitative differences in tongue length and width between alveolar and velar plosives. If there are any quantitative differences in tongue shape, length and width are present, are they statistically significant?
3. Whether there are any quantitative differences in tongue length and width between velar plosives and velar nasal stops and whether there are any differences between /k/ and /ɡ/. If any quantitative difference in tongue shape, length and width are present, are they statistically significant?
4. Whether there was a change in tongue shape post-therapy, in each of the three comparisons (VAM, UVBF and BL-M).

Single words from the untreated velar wordlist were annotated. As was done with Andrew, the plosives /k/ and /ɡ/ tokens were annotated at the burst whereas a midpoint was selected from /ŋ/ tokens. Corresponding ultrasound frames were then selected and splines indicating the tongue surface were fitted to the images using the semi-automatic edge-detection function in AAA. Splines for /k/, /ɡ/ and /ŋ/ were then averaged and compared within session in the AAA workspace. See Figure 56 and Figure 57 for examples of multiple tokens and an averaged tongue spline for /k/. As there was no visible palate trace to align the data across sessions, within-session comparisons were made and longitudinal comparisons are on the basis of these extracted measures.

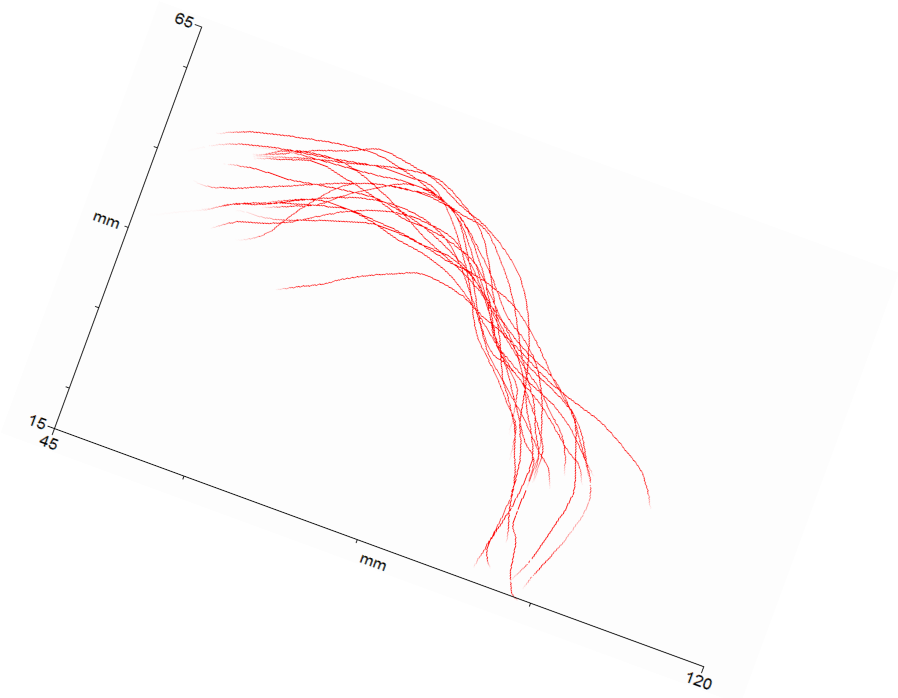


Figure 56 Example of multiple tokens of /k/ (rotated by 20°)

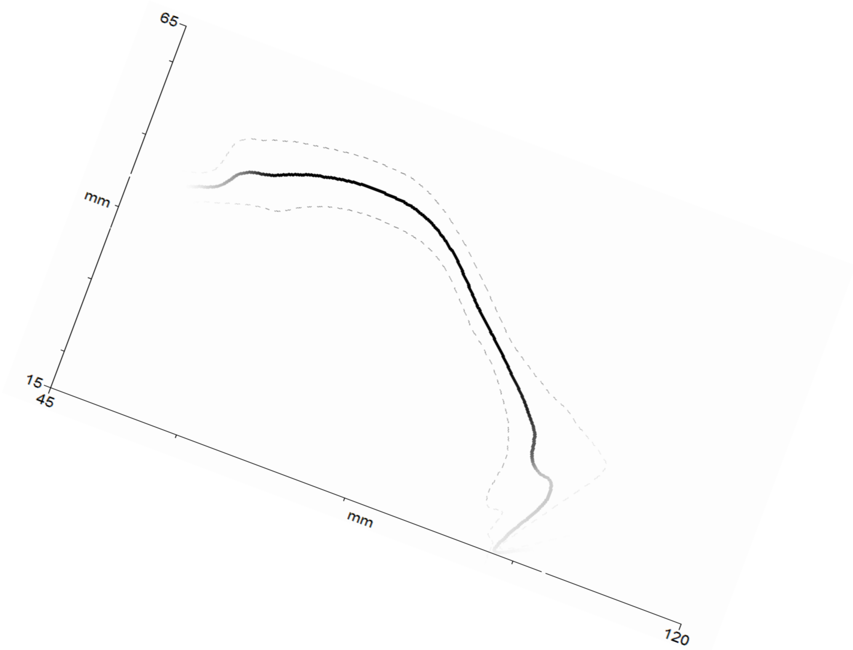


Figure 57 Example of averaged /k/ from the multiple tokens in Figure 56 (rotated by 20°)

Similar to Andrew's data, two quantitative measures of the similarities/differences of two tongue shapes are used to decipher the maximum and mean width and the length of the tongue for /k/, /g/ and /ŋ/, and significant differences are also reported and quantified. As Craig's productions on velar nasals were transcribed as correct, comparisons were made between /k/ and /ŋ/ and /ɡ/ and /ŋ/ to determine how similar or different his productions of velar plosives were to velar nasals pre- and post-therapy and whether this changes longitudinally. A comparison of /k/ and /ɡ/ was also measured, to determine any differences in voiced and voiceless productions, on the one hand because these were transcribed differently and also because /ɡ/ was targeted more frequently in therapy than /k/. Phonetic transcriptions showed that pre-therapy, Craig's productions of /ɡ/ were realised as either [d] or [n], whereas /k/ was produced mostly as glottal stop. Therefore, the two differently voiced targets might behave differently. Quantitative measures will be able to determine whether there were covert errors, such as double articulations, and whether there was lingual movement in the velar or alveolar region.

Secondly, all 50 items from the DEAP were annotated. Where present, i.e. where Craig produced a substitution and not an omission, all tokens of /t/ (N=12) /d/ (N=3), /k/ (N=12) and /ɡ/ (N=5) were annotated at the burst with corresponding key frames splined. Splines were averaged for alveolar plosives /t/ and /d/ (N=15), and velar plosives /k/ and /ɡ/ (N=17) and compared within session. Quantitative measures were carried out to look at the similarities and differences in alveolar and velar stops in the DEAP. As the therapy target changed in therapy block two to alveolar /t/, but there was no baseline for an untreated wordlist, the DEAP was used to identify any changes in alveolar tongue shape throughout therapy. Another rationale for comparing the alveolar and velar tokens from the DEAP, although from different vowel environments, is the lack of minimal pairs in the wordlists for Craig. Whilst an additional wordlist was designed with distractors and minimal pairs, this was not in fact recorded, due to Craig's motivation and concentration levels not being sufficient to undertake the task.

4.1.2.4 Ultrasound Image Quality and comparisons to TD tongue shapes

In addition to the within-speaker evaluation, both Andrew and Craig's ultrasound images are compared to age-matched typically developing speakers. A qualitative comparison of the quality of the raw images will be presented, along with a comparison of tongue length for their age-matched peers. This will be discussed throughout the results and will not be presented separately.

In addition, qualitative interpretations of the tongue shapes in terms of traditional phonetic categories of place will be presented, and compared to the similar interpretations which are implicit in the transcriptions.

The following sections will present the qualitative and quantitative results and then consider the implications of the results, including the difficulties in recording and analysing ultrasound data with children, particularly those with cleft palate with or without syndromes.

4.2 Articulatory Results and Discussion

4.2.1 Andrew

Qualitative and quantitative measures showing the contrast between /n/ (therapy target) and /ŋ/ (for comparison to the most common substitution for /n/), through analysis of minimal pairs will be presented first, followed by a qualitative analysis of /n/ vs. /t/ and /n/ vs. /k/ in /i/ /o/ and /a/ in CV or CVC single tokens and a qualitative comparison of /n/ in different word positions. Lastly, measures from the DEAP will be presented, with qualitative comparisons of /t/, /k/, /n/, /ŋ/, /s/ and /ʃ/, and quantitative measures of /t/ vs. /n/ and /k/ vs. /n/.

4.2.1.1 Contrast between /n/ and /ŋ/ (minimal pairs)

Perhaps the most important result to come from the quantitative articulatory analysis is that, contrary to the phonetic transcriptions identifying no contrast between /n/ and /ŋ/, qualitative and quantitative analysis of the ultrasound data identified a covert contrast between /n/ /ŋ/ in the well-controlled minimal pair wordlist.

All minimal pair tokens were transcribed as [ŋ], therefore an identical tongue shape would be expected for /n/ and /ŋ/, appropriate for /ŋ/ of course, but not for /n/.

However, Figure 58 shows a clear difference in Andrew's productions of /n/ (green) and /ŋ/ (purple). As the +/- one standard deviations in Figure 58 do not overlap, this gives a visual clue that there is likely to be statistical significance (radially) between the means within the non-overlapping area, much like the visual interpretation of non-overlapping confidence areas in Smoothed Spline ANOVA (SS-ANOVA) diagrams (Davidson 2006).

As it was not possible to trace the palate or use a bite plane, a +20° rotation (clockwise, with anterior to the left) was adopted, similar to that in Scobbie et al. (2011) and Scobbie et al. (2012), to help interpret the place of articulation involved and to determine, for example, what the actual phonetic realisation of /n/ was. And, while it might seem reasonable to assume that the correctly transcribed /ŋ/ is velar, in the ultrasound data, this too was evaluated.

This comparison provides another surprising result. Andrew's tongue shape for /n/ is in fact retracted *further* than /ŋ/, with tongue root elevation potentially indicating some uvular contact. This is likely to be a residual compensatory articulation due to ongoing VPD, with retraction to uvular placement being a common compensatory articulation in speakers with CP. /ŋ/ images show elevation in the dorsal region, as expected, with some possible whole-tongue body gestures, reported in Gibbon (2004)'s EPG paper as involving increased contact between the tongue and hard palate. To confirm whether whole-tongue body gestures were present in the ultrasound data, a palate trace would be required. However, due to the quality of the data, it was not possible to trace the hard palate on Andrew's data.

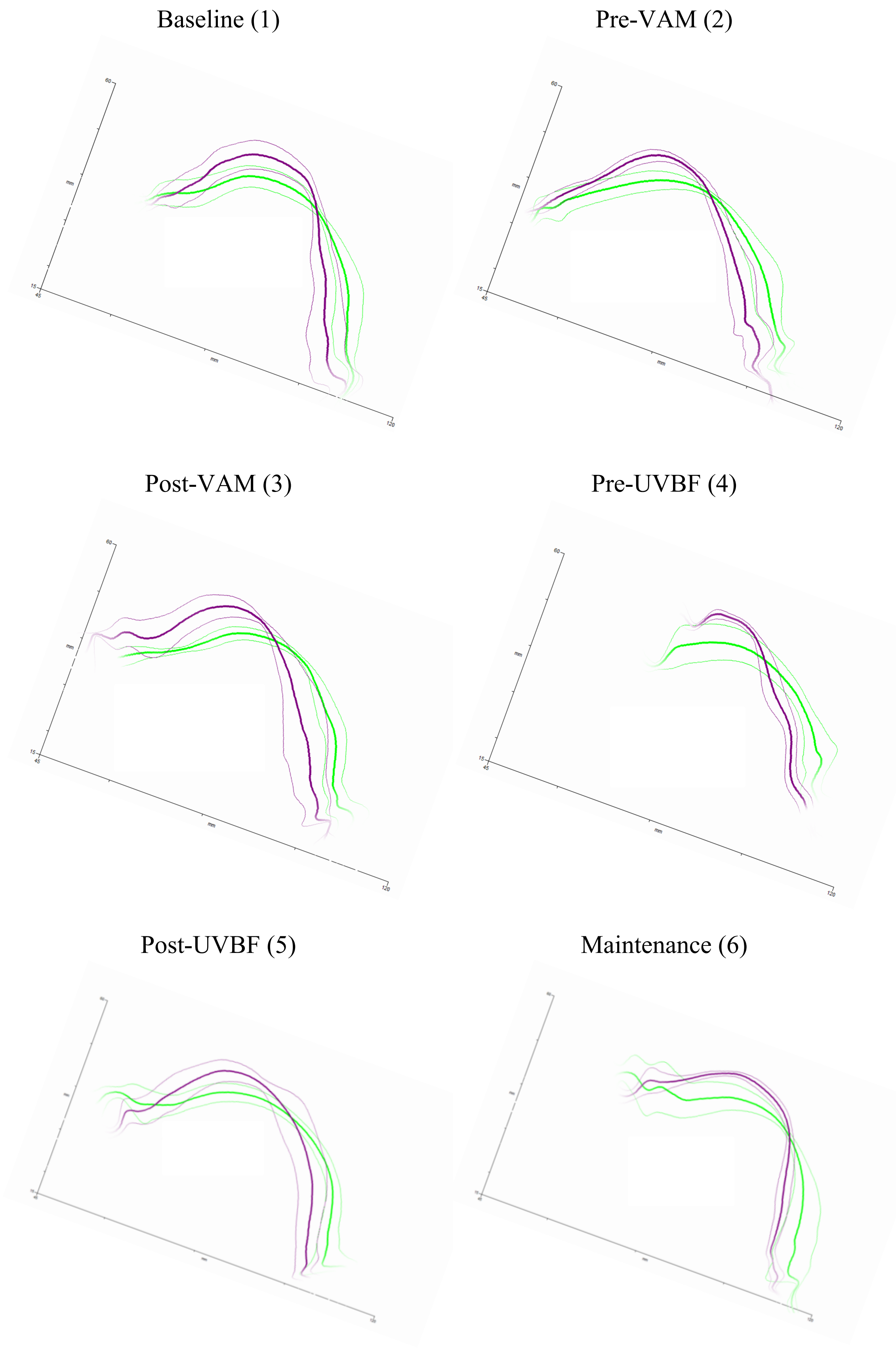


Figure 58 /n/ (green) /ŋ/ (purple) minimal pairs in assessment sessions one to six

It is evident from the images in Figure 58 that there is some data missing in session four (pre-UVBF), thus suggesting that the probe was angled incorrectly in this session. With qualitative analysis showing a clear contrast between Andrew's productions of /n/ and /ŋ/, the next step was to quantify these differences.

Figure 59 presents a range of information which can be used to quantify the size of the difference between the /n/ and /ŋ/. First it is necessary to consider those parts of /n/ and /ŋ/ which were tracked with enough confidence to be quantified. A confidence threshold of 80% was used across sessions – this is an internal (i.e. arbitrary) value assigned by AAA during its spline-fitting, and was selected on the basis of a qualitative examination of the data. Mostly a spline is fitted with 100% confidence, but where the bright white areas of the image created by the tongue surface disappear at the anterior and posterior limits of the data, there is a short transition as the confidence drops to 0% and the spline ends. Lower confidence areas are also possible where the distinctness of the edge is less obvious for other reasons. When the individual splines are averaged together, therefore, a composite confidence value for each radial point defining the mean spline is obtained. The length of confidently-splined tongue surface therefore is a measure of "what is available" for subsequent analysis.

Figure 59 shows the "total" visible length for /n/ and /ŋ/ drawn from the set of minimal pairs. Where both /n/ and /ŋ/ had a confidence of over 80% for the same radial point, is it possible to test for a significant distance from the probe to the tongue surface statistically, by t-test, as described above. To provide a measure of how much comparison is possible, the comparable lengths for /n/ and /ŋ/ are then given. The maximum comparable length (63mm) was found for /n/ in session five (post-UVBF), with the minimum length overall (33mm) found in session four (pre-UVBF) for the comparable length of /ŋ/. This is unsurprising, given the poor image quality for session four (Figure 58).

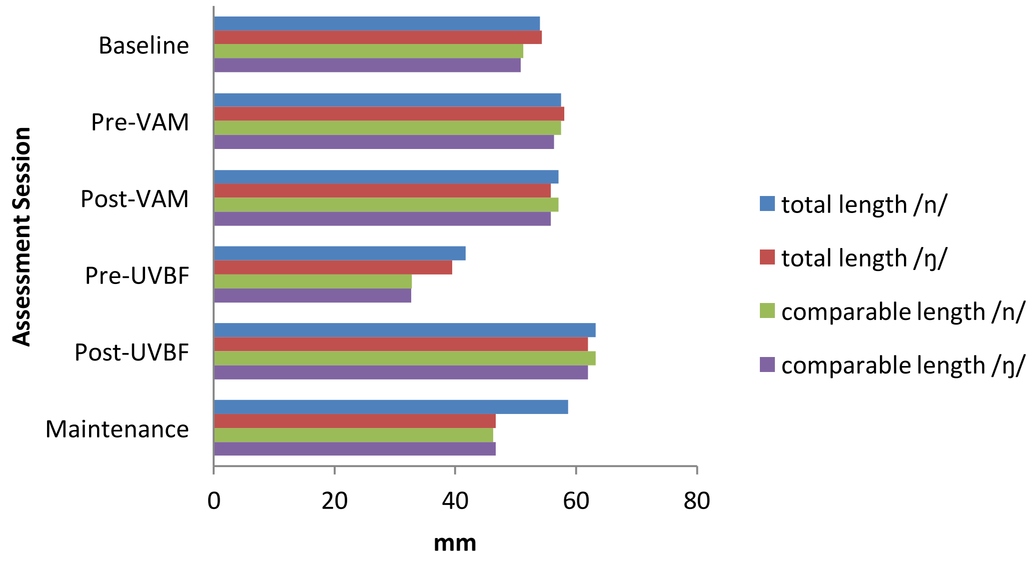


Figure 59 Andrew's Tongue Length Visible for /n/ and /ŋ/ in minimal pairs

To compare these lengths to Andrew's age matched peers, it would be expected that the average total visible length for /t/ in typically developing children is 62mm at age 9;0 and 65mm at age 10;0. Only the post-UVBF recordings have a visible tongue length that fits within this age-matched range, perhaps suggesting that his actual physiological tongue length is less than the normal range for his age. However, this may also indicate that the quality of Andrew's images in the other five sessions is also somewhat poorer than in the images of his typically developing peers, with the length of visible tongue being shorter than expected in these sessions.

Turning now to the more linguistically and clinically important results that convey how different the targets are, paired t-tests showed significant differences between these two targets in all six sessions along the tongue length. Of course, not all the tongue surface needs to be different, though there is a minimum threshold of five contiguous radial tests (see above). However, it is interesting that such a high proportion of the comparable tongue length, for both /n/ and /ŋ/, was identified as statistically significantly different (Figure 60). In every session, over 80% of the comparable tongue length was signficantly different. The pre-UVBF session had the highest proportion of significantly different tongue (100% for both /n/ and /ŋ/). However, this is likely to be due to the poor image quality which as we have seen reduced the comparable visible tongue length in that session. The lowest proportion of significance was found in the post-UVBF session (83% for /n/ and 82% for /ŋ/).

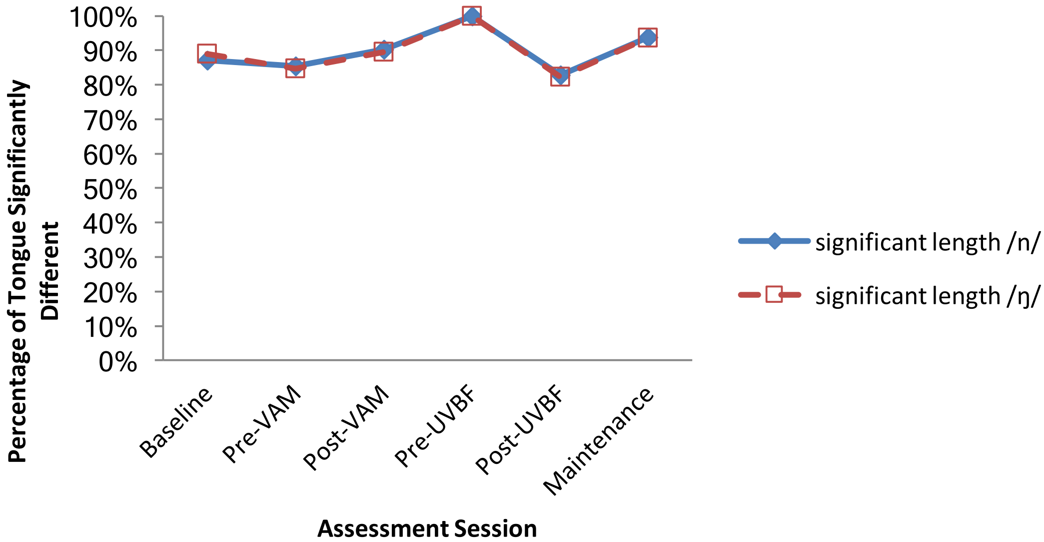


Figure 60 Proportion of the Tongue Length Identified as Being Significantly Different in Minimal Pairs

The other important quantitative measure of difference between the two targets reports how far apart they are. The mean width (difference in probe-to-surface distance along each fan line) between /n/ and /ŋ/ is charted in Figure 61. The figure shows both the whole comparable tongue length (red), which will therefore include parts of the tongue surface where the absolute difference between the targets is not actually statistically significant (and therefore is small), and the average width within the somewhat shorter length of the tongue that is significantly different (if greater than five adjacent fan lines, and additionally including any cross-over (as defined above)). The mean width difference between /n/ and /ŋ/ along the "significant zone" was largest post-VAM (5.2mm), with the baseline having the smallest value (3.5mm), indicating that post-VAM has the widest average significant difference in tongue shapes. Though there is no direct analysis of the significance of these session comparisons, the descriptive results are in-line with phonetic transcriptions that showed the largest increase in PTCC was post-VAM. The width difference along the entire comparable tongue length was also largest in post-VAM (5.0mm) and smallest in baseline (3.3mm).

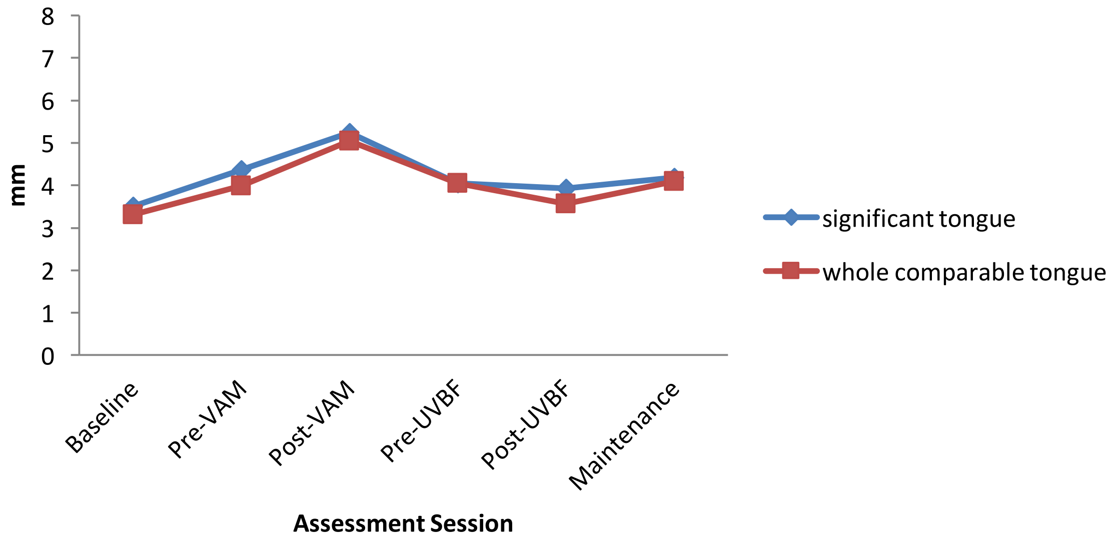


Figure 61 Mean width between /n/ and /ŋ/

Finally, it is useful to consider not just the mean width of the difference, but the maximum single radial difference found in a given session long the comparable tongue length (Figure 62). These values are of course larger. The longitudinal pattern is similar again, with the maximum difference between /n/ and /ŋ/ also found in post-VAM (7.4mm), with the lowest in baseline (5.3mm).

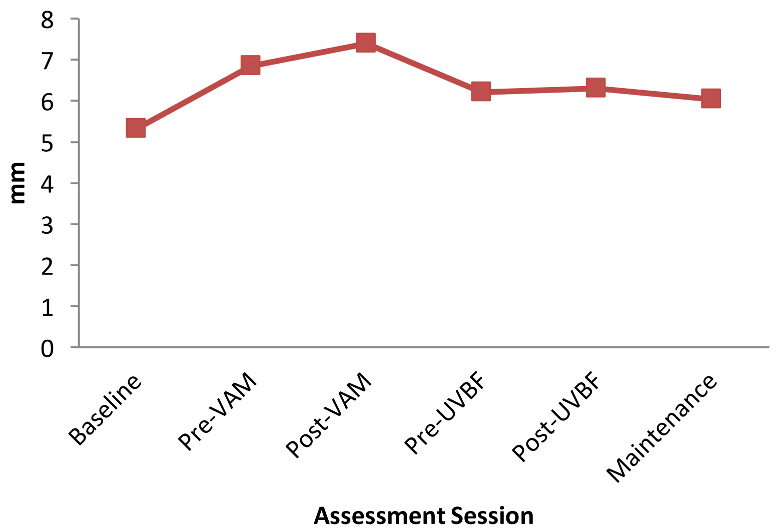


Figure 62 Maximum width between /n/ and /ŋ/ for the whole comparable tongue length

4.2.1.2 Comparison of /n/ to /t/ and /k/ (additional alveolar wordlist)

In order to look at how phonetically alveolar or velar the phonologically alveolar /n/ is at a single-token level, qualitative analysis was carried out to compare single tokens of /n/ to a phonologically alveolar comparison segment /t/ and a phonologically velar /k/, both of which appear phonetically accurate. The comparisons were made in WI position in three separate vowel environments (/i/, /o/ and /a/). These tokens were taken from the additional alveolar wordlist, where there were only single tokens of WI CV or CVC. In typical speakers, it would be expected that /n/ and /t/ are phonetically the same (Gibbon et al. 2007) with tongue tip raising and the tongue dorsum lowering. On the other hand, /k/ is expected to have tongue tip lowering and dorsum elevation. There will be coarticulatory differences depending on vowel environment. In an /i/ environment, it would be expected that /k/ would be more palatalised. Andrew's productions of /n/ were transcribed as [ŋ] in all three vowel environments, indicating that a tongue-shape closer /k/ than to /t/ would be expected.

In general, Andrew's /n/ does not have raising of the tongue blade towards the alveolar ridge to the extent seen in /t/, except in the maintenance session. In other sessions, though tip raising for /n/ is visible, it usually is visibly less than the amount of tip raising for /t/ and may be similar to the position observed for /k/. Recall that ultrasound will not image the tongue tip if the beams have to pass through a sub-lingual airspace, which is typical for alveolar contact, but while care has to be taken in interpreting the images, the comparison between tongue curves can be based on articulatory similarity or difference together with some key assumptions, such as the presence of tongue tip contact for /t/ and its absence for /k/. Some specific exceptions are noted below.

Figure 63 shows /n/ (green), /t/ (red) and /k/ (blue) in ‘nap’, ‘tap’ and ‘cap’ in all six assessment sessions. Within the first four assessment sessions, the tongue root for /n/ is retracted further than both /t/ and /k/, however less so in the post-UVBF and the maintenance sessions. Post-UVBF shows a double articulation for /n/, similar to that in ‘know’ (see below). In the post-UVBF session, /k/ also appears to be double articulated, with the anterior portion of the tongue in /n/ being in a similar zone to that of /t/. The anterior portion of the tongue looks reasonably correct in the maintenance session; however, there remains a similarity in the tongue root for /n/ and /k/. There also appears to be some retraction of the tongue root for /t/ in maintenance. Given that it is expected that correct /t/ and correct /n/ are the same, therefore the data here shows that /n/ is generally retracted, and for a child with a submucosal cleft, a likely interpretation is that this has been a carry-over of a previous cleft-type characteristic (or a residual incorrect motor plan). As all tokens of /n/ were transcribed as velar, these images are not surprising. Double articulations appear more frequently in post-therapy sessions, indicating that the tongue tip raising is occurring, in those sessions, which is interpreted as a result of therapy.

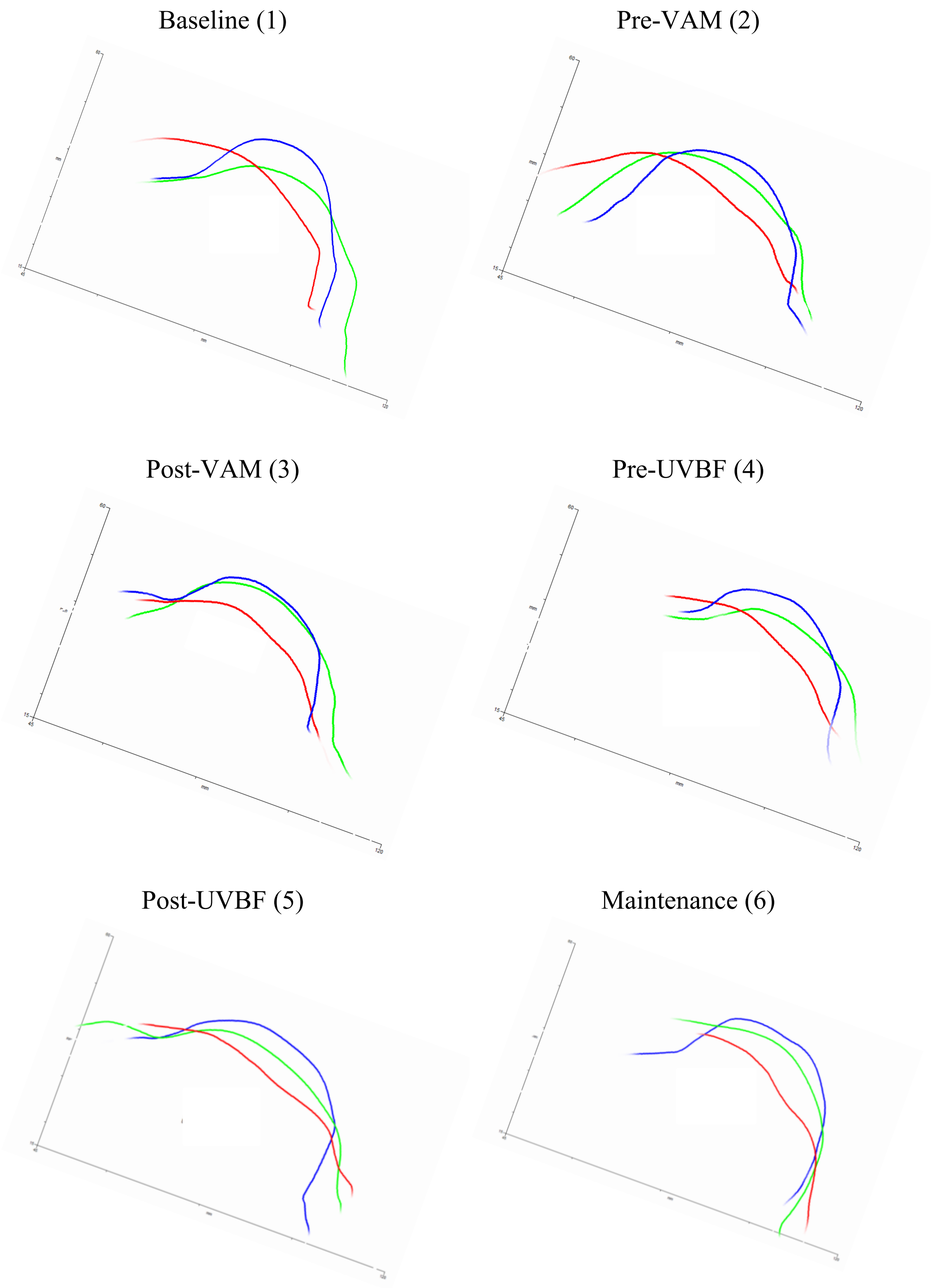


Figure 63 /n/ (green), /t/ (red), and /k/ (blue) in 'nap', 'tap' and 'cap' in all six assessment time-points

As well as double articulations for /n/, there were also indications that the whole tongue body was raising for /n/ and for /k/ (which is inappropriate, and appropriate, respectively).

It is possible to draw on some other evidence to clarify these ambiguities. In Figure 64, for example, it is not clear if the tongue tip is raising in the ultrasound image for ‘cap’ in the post-UVBF session, due to a mandible shadow, although it does indeed look as though the tongue tip is extending forward. However, the elevated tongue tip is in fact visible on the lip video data. In this case, the lip camera data is providing additional information to the ultrasound data. In other cases, the visible tongue tip on the lip camera data confirms the interpretation of the ultrasound images given above – the combination of UTI and lip camera supports the judgement that the very extreme tongue tip itself is elevated.

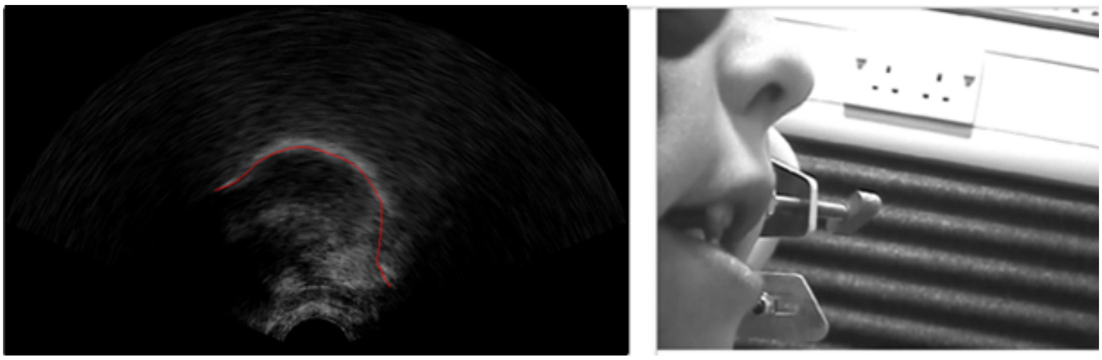


Figure 64 Example of whole tongue body gesture or double articulation for /k/ in 'cap' post-UVBF

Figure 65 (below) shows /n/ (green), /t/ (red) and /k/ (blue) in ‘know’, ‘toe’ and ‘co’ in all six assessment sessions. In the pre-VAM and pre-UVBF sessions, /n/ and /k/ are similar shapes. Due to co-articulatory effect, it would be expected that all tongue shapes of these consonants before /o/ would be more posterior than in an /i/ environment where they may be expected to be fronted/raised more toward the palatal region. At baseline, the tongue shape for /n/ is flatter than /k/, however /n/ still shows a retracted placement. In post-VAM, the tongue shape for /n/ is more similar to /t/, however is slightly higher. The tongue root in /n/ is retracted further than both /t/ and /k/ in post-UVBF, with a cross over in tongue shapes for /t/, /k/ and /n/ in the palatal region. In the maintenance session, placement for /n/ is in the post-alveolar or palatal region with bunching in the dorsal region of the tongue. It is possible that there is alveolar contact in maintenance but there is little evidence in the image of this. However, in the other sessions, there is either an indication that /n/ is not alveolar, by being different to /t/, or that /n/ /t/ and /k/ are all have similar tongue surface locations in the anterior part of the image. Maintenance is the only session in which /n/ and /k/ look distinct. This suggests that although PTCC scores increased post-VAM indicating improvement in alveolar placement, perhaps only with ultrasound a firm articulatory underpinning and alveolar/velar contrast was possible.

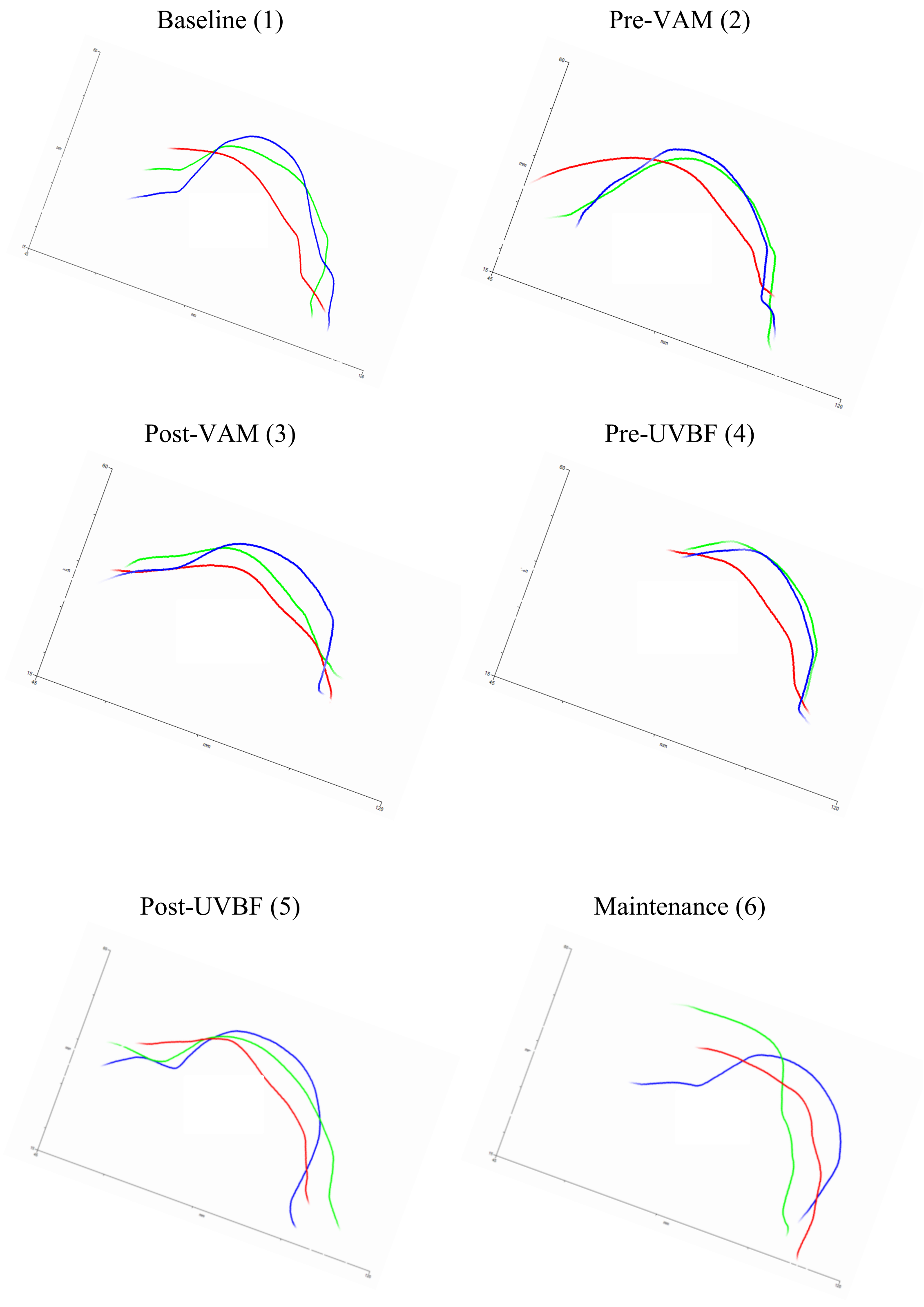


Figure 65 /n/ (green), /t/ (red), and /k/ (blue) in 'know', 'toe' and 'co' in all six assessment time-points

Figure 66 (below) shows /n/ (green), /t/ (red) and /k/ (blue) in ‘knee’, ‘tea’ and ‘key’ in all six assessment sessions. As noted above, there is no clear alveolar contact for /n/, except in the maintenance session where there is a clear indication of elevation toward the alveolar ridge. However, as there is no palate trace, it is difficult to tell exactly where the tongue tip is in relation to the alveolar ridge.

At baseline, /n/ shows similarities to /k/ with the tongue tip lowered and the dorsum raised, with slightly more anterior contact than in /k/. Although Gibbon et al. (2007) suggest that in typical adults alveolar oral and nasal stops are the same, it is unknown if this is also the case for velars. Therefore, it cannot be certain from these qualitative analyses that even when /n/ appears to be similar to /k/ that this is suggestive of a merger with /ŋ/. In fact, the data from the minimal pairs above would suggest otherwise, with an identified covert contrast between /n/ and /ŋ/.

In the pre-VAM session and more obviously in the post-VAM, palatal contact for /n/ looks slightly more retracted than for /k/, with tongue tip lowering in both /n/ and /k/ and raising for /t/. In the post-VAM session, the tongue root for /n/ is higher than both /t/ and /k/. In the pre-UVBF session, there was a large mandible shadow, therefore a lot of the tongue-tip data is missing. However, tongue shapes for /n/ and /k/ are near-identical, indicating a phonetically velar place of articulation for the phonologically alveolar nasal. Post-UVBF, there is elevation of the tongue-tip toward the alveolar region for all three consonants. Whilst it appears that there is a whole-tongue body gesture for /k/, /n/ has a different presentation, with a dip in the tongue, suggestive of an alveolar/velar double articulation. Also in the post-UVBF session, /t/ appears to have a secondary place of articulation in the tongue root region. Audible double articulations are frequently reported in CP literature, and a strong confirmation of these atypical errors along with covert ones exists in the EPG literature (Gibbon 2004). In the maintenance session, /n/ is more anterior than /k/ with tongue-tip elevation; however, the tongue dorsum is not as low as in /t/.

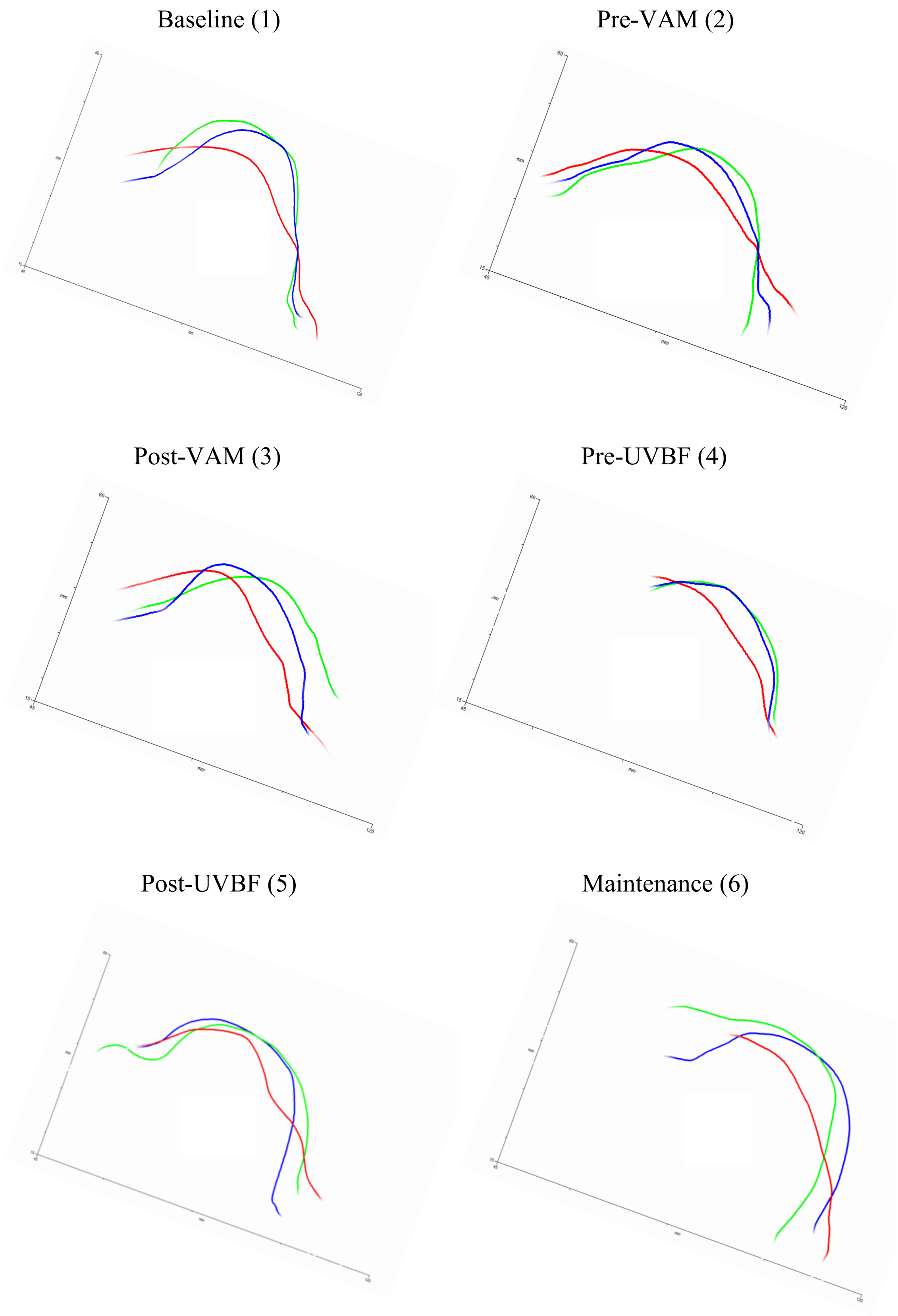


Figure 66 /n/ (green), /t/ (red), and /k/ (blue) in 'knee', 'tea' and 'key' in all six assessment time-points

4.2.1.3 Tongue shape for /n/ in different word positions in Untreated /n/

Figure 67 shows the ultrasound tongue shapes for word initial, word medial and word final position in the untreated /n/ wordlist for each of the six assessment sessions. Although the probe was in a different orientation within each session, all six frames show a similar session-internal tongue pattern for /n/. There are no qualitative differences between the average tongue shapes in word initial (orange), medial (green) or final position (purple), indicating that word position has no consistent effect on the tongue placement in Andrew's production of /n/ from session to session. The lack of any word position result also helps reinforce the value of the significant results found earlier when contrasting /n/ to other segmental targets.

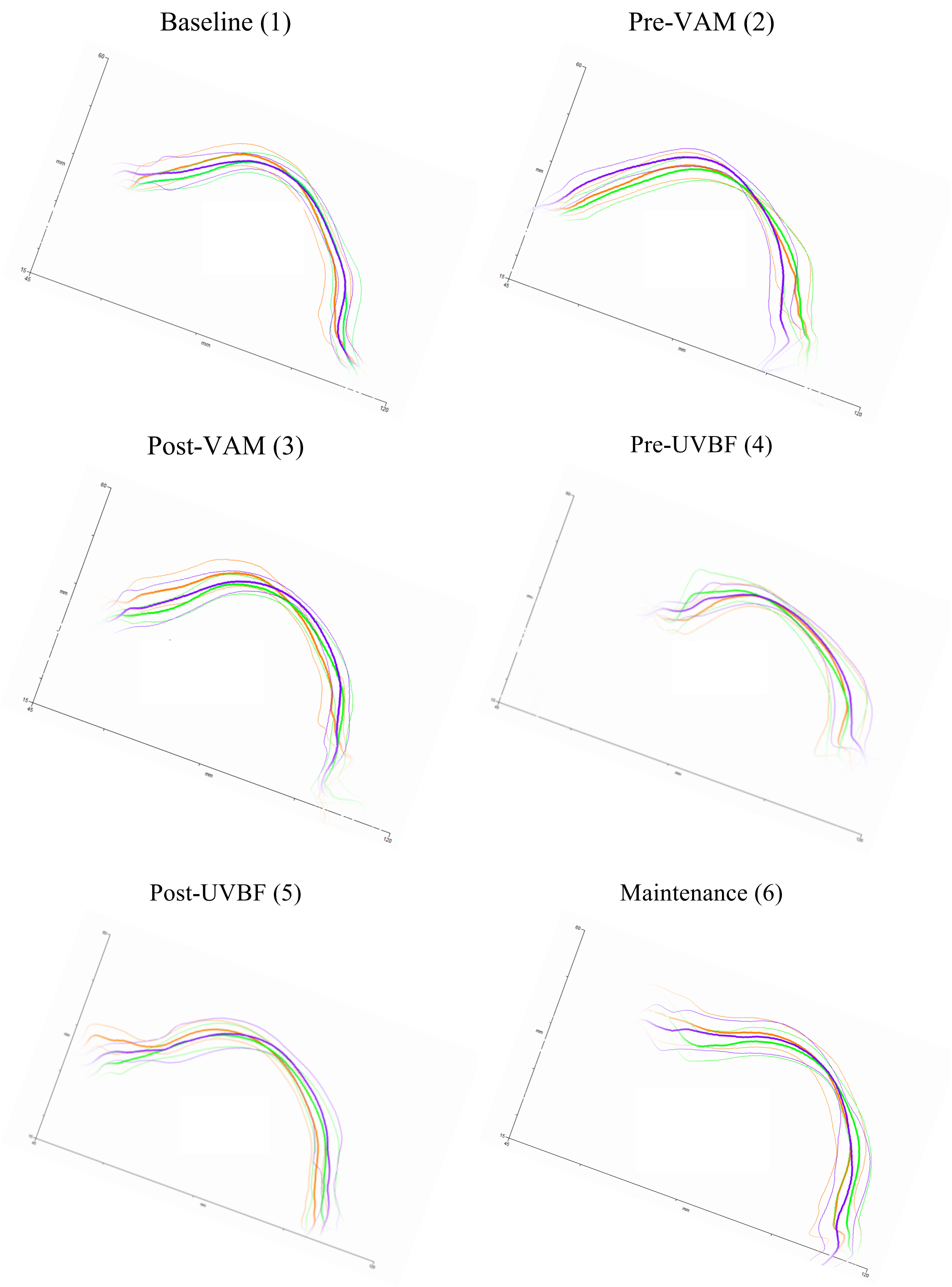


Figure 67 Untreated /n/: Comparison of word positions in assessment sessions one to six

(orange = WI, green = WM, purple = WF)

4.2.1.4 Analysis of single words from the DEAP

In order to further explore the articulation of /n/, qualitative measures were carried out to explore the wider consonantal space. This was enabled by analysing tongue shapes from a range of alveolar, post-alveolar and velar tokens within the DEAP phonology subtest. /n/ was qualitatively compared to other alveolar tokens (/t/ and /s/) and /ŋ/ to other velar tokens (/k/). /s/ and /ʃ/ were also included to show examples of post-alveolar placement to show the space between alveolar and velar. The DEAP does not, unfortunately, contain any palatal consonants, such as the voiceless palatal fricative allophone of /h/ in "**h**uge". Table 57 gives an overview of the number of tokens for each consonant included in the average tongue splines. This number is not equal across consonants due to the imbalanced number of tokens included in the DEAP phonology subtest.

Phoneme	Number of tokens
/t/	12
/k/	12
/s/	14
/ʃ/	4
/n/	8
/ŋ/	4

Table 57 Number of tokens included in the averages for tokens in the DEAP

Figure 68 shows the average tongue shapes for /t/ (red), /k/ (blue), /s/ (turquoise), /ʃ/ (pink), /n/ (green) and /ŋ/ (purple). A similar trend can be seen in all six sessions for /t/ and /s/, which are near-identical in shape, showing alveolar placement but little indication of the manner difference. This alveolar similarity fits with the phonetic theories that entail the notion of primary place of articulation (Ladefoged and Maddieson 1996). However, the tongue *body* positions are also close in all sessions. The tongue blade of /ʃ/ is consistently close to /t/ and /s/ but is more bunched, showing post-alveolar placement. In all sessions, its dorsum is, unsurprisingly, lower than /k/ and /ŋ/, but also it is lower than /n/, reflecting the latter's atypical production. Note, however in maintenance there is more palatal constriction for /ʃ/

than in other sessions and it stands out more from the other consonants. As expected, /k/ and /ŋ/ show similar tongue shapes throughout, with more dorsal raising for /ŋ/, particularly in the pre-UVBF session, apart from in the maintenance session where they are near-identical.

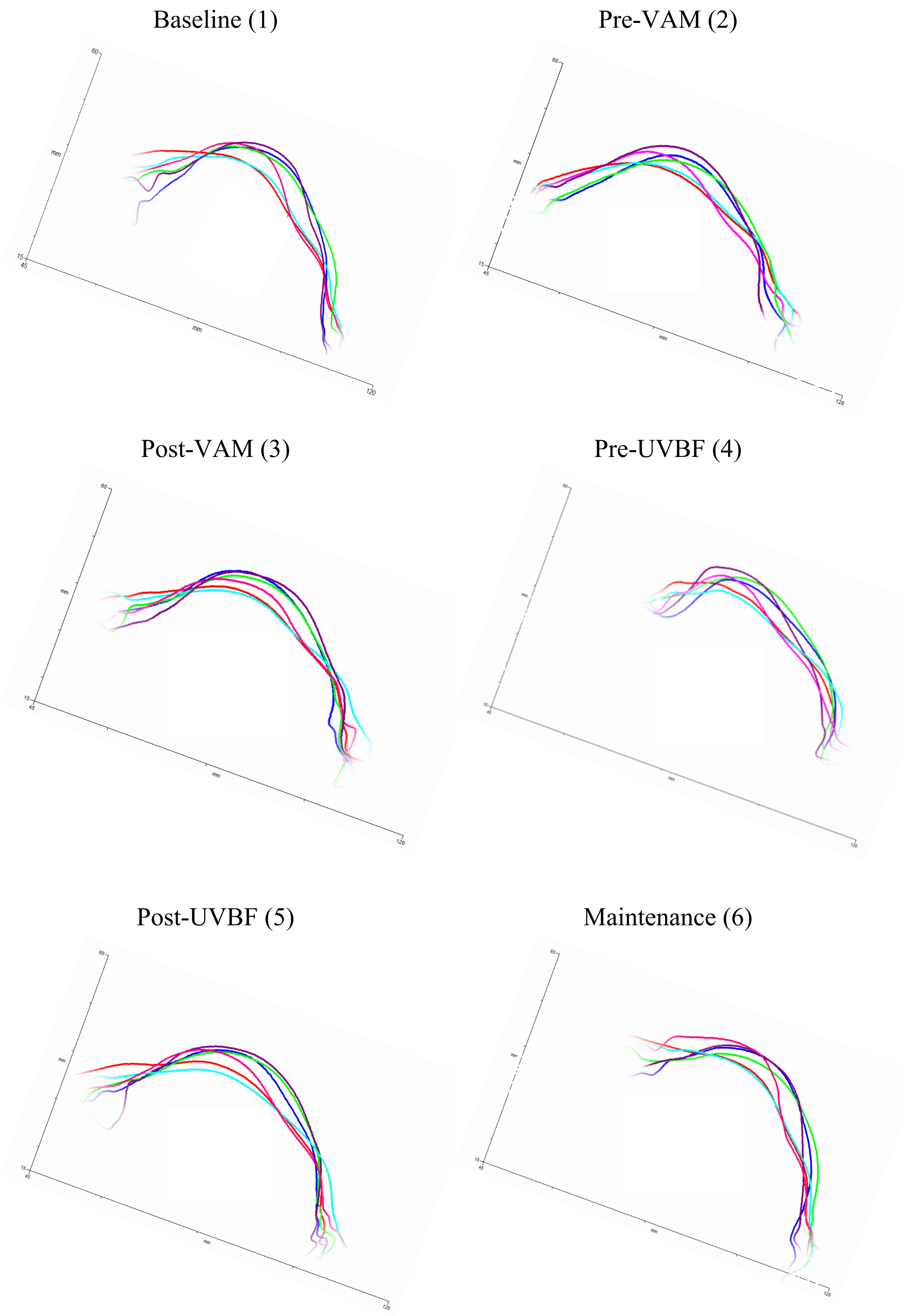


Figure 68 DEAP: Averages for /t/ (red) /k/ (blue) /n/ (green) /ŋ/ (purple) /s/ (turquoise) /ʃ/ (pink) in all six assessment sessions

Finally, it is possible to consider the relative placement of /n/ compared to these other consonants. Similar to the /t/, /k/ and /n/ comparisons in minimal pair data, presented above, there is a clear contrast between /t/ and /n/, when these would be expected to be similar in shape in typical development, whereas /k/ and /n/ show similarities that ought to be absent. And when comparing the two nasals, /n/ and /ŋ/ show more similarities in the DEAP (Dodd et al. 2002) data than in the minimal pair data.

A drawback of presenting averaged data as a mean shape is that a raw mean does not indicate the degree or nature of the variation in the underlying tokens. Even when including the standard deviation, averaging tongue shapes does not, however, show if there are random or continuous variation in shape within each consonant, or something more patterned, like a bimodal variation between two alternatives.

Consider, for example, Andrew's production of /k/, which, because it was mostly correct in simple words in transcription, has been represented above as a single mean trace as a velar point of comparison for the backed /n/. Within each session, there was a large variation within phonetically more alveolar and more velar tokens of /k/ in more complex words. For example, in the pre-UVBF session Andrew had a range of velar plosives which were transcribed as velar, alveolar, palatal or a double articulation. Figure 69 shows the variation in velar productions in the DEAP which can explain these transcriptions. First there are three outliers which may indicate headset movement or deletion of the target. When these three are omitted, there are two clear categories – one appears to be a velar placement with raising of the tongue dorsum and lowering of the tip (denoted in Figure 70 as blue), and the other is a bunched shape with elevation of the tongue tip to post/alveolar or palatal region and lowering of the dorsum (denoted in Figure 70 as red).

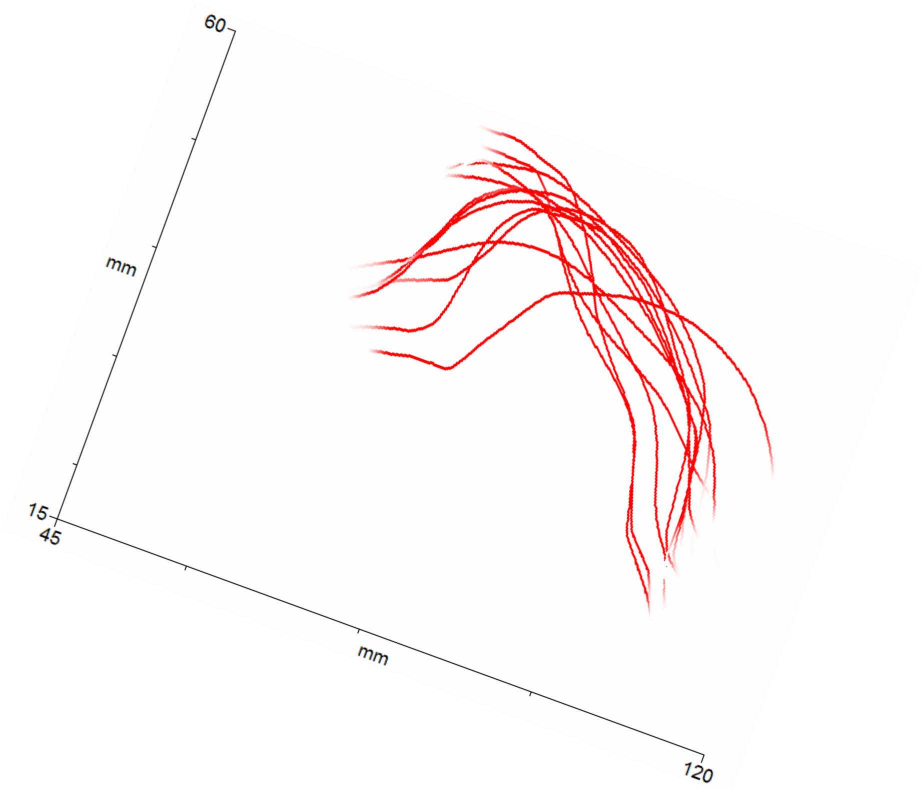


Figure 69 Variation in production of /k/ in the DEAP pre-UVBF

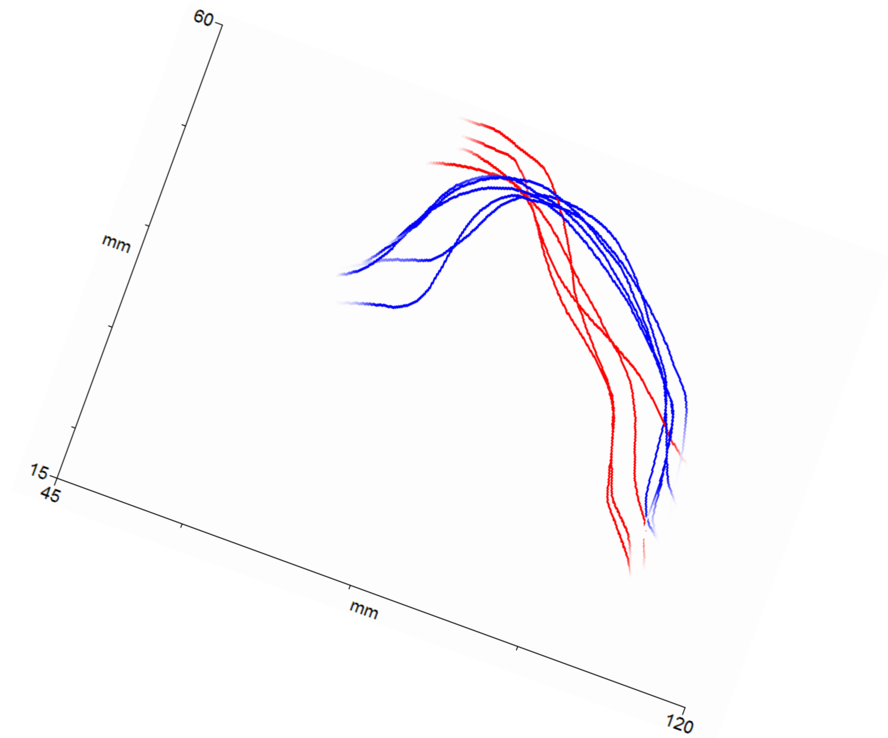


Figure 70 Variation in production of /k/ separated into two categories (with three outliers omitted)

Another interesting finding in the DEAP was when there were contrasts identified in transcriptions but not in tongue shapes. For example, in the baseline session, the velar plosive /g/ was transcribed differently in egg ([ɛɟ]) and pig ([pɪg]) by two phonetically trained listeners (including the tSLT), however, when looking at the images they are near identical in palatal placement (Figure 71).

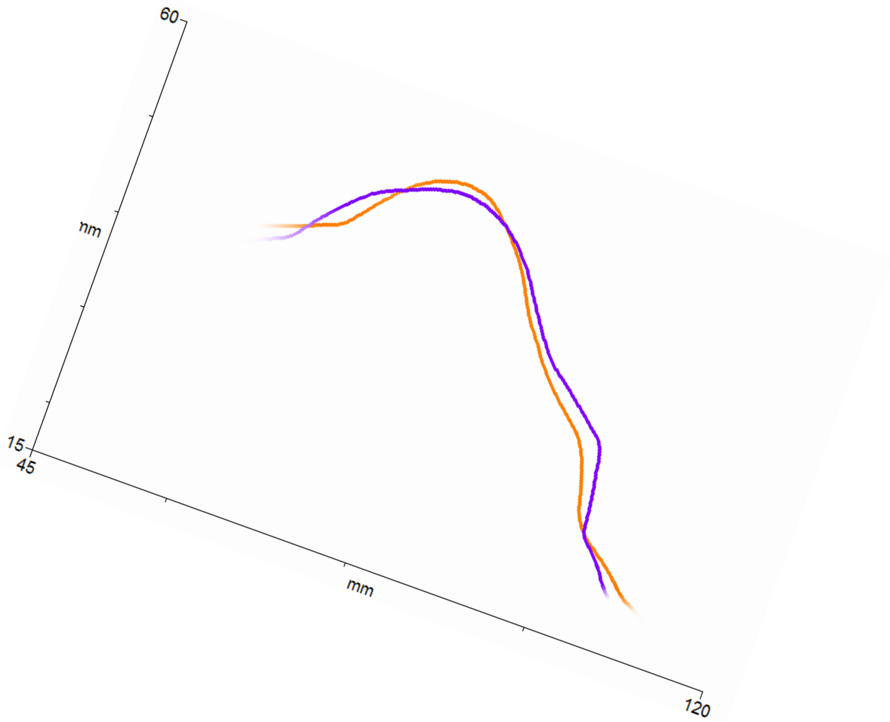


Figure 71 Tongue shapes for /g/ in pig (purple) and egg (orange) at baseline

In order to further quantify how alveolar or velar-like /n/ is, /n/ is quantitatively compared to /t/ and /k/ separately. Statistical analysis using the built-in t-test function in the AAA software showed significant differences between /t/ and /n/ from the DEAP data in all six sessions. On the other hand, there was no difference found between /k/ and /n/ in any of the six sessions. Quantitative measures for /n/ vs. /t/ and /n/ vs. /k/ are presented separately below.

4.2.1.5 Quantitative analysis of /n/ vs. /t/

Figure 72 shows the total length for /n/ and /t/ from the DEAP. Where both /n/ and /t/ had a confidence of over 80%, this is shown as the comparable length for /n/ and /t/, as above. The maximum length (62mm) was found in the total length for /t/ in session two (pre-VAM), with the minimum length (33mm) found in session four (pre-UVBF) for the comparable length of /t/, similar to the minimum length in the minimal pairs data for pre-UVBF. Again, when comparing these lengths to Andrew's age matched peers, it would indicate that the tongue length for /t/ and /n/ is below the normal range for his age, with poor image quality found in five out of six sessions.

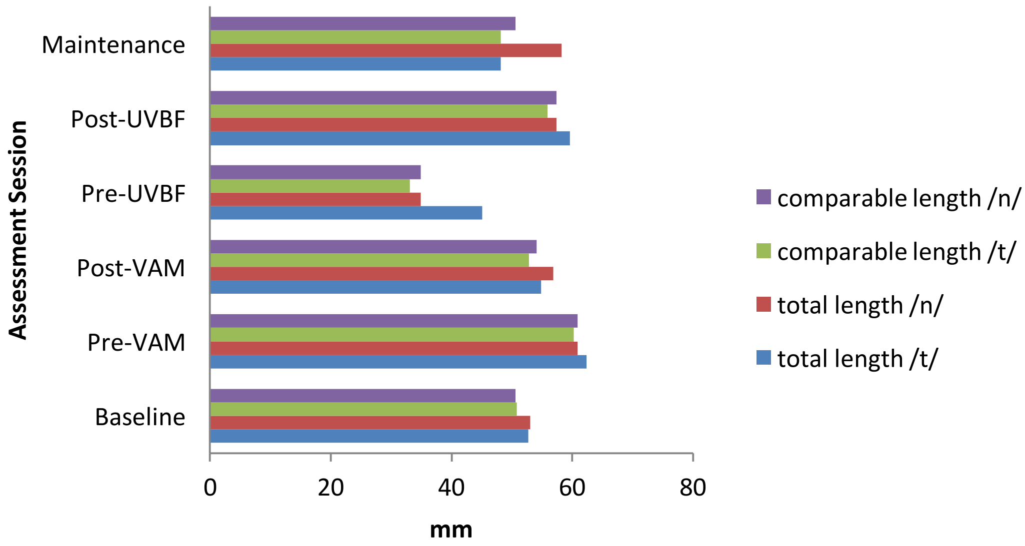


Figure 72 Andrew's Tongue Length Visible for /n/ and /t/ in the DEAP phonology subtest

The proportion of the tongue that is significantly different was calculated using the paired t-test function in AAA (see Figure 73). As the phonetic transcriptions showed that /t/ was transcribed as correct and /n/ was transcribed as being velar, a statistically significant difference between /t/ and /n/ was expected. All sessions showed over 60% significant difference across the comparable tongue. The lowest proportion of significance was found in pre-UVBF, which is also the session where there was the highest proportion of significant difference between /n/ and /ŋ/ in the minimal pairs, suggesting that /n/ was perhaps closer to the alveolar target in the pre-UVBF session. However, this could also be due to the poor image quality in this particular session, with a large part of the image of the anterior portion of the tongue missing from the data.

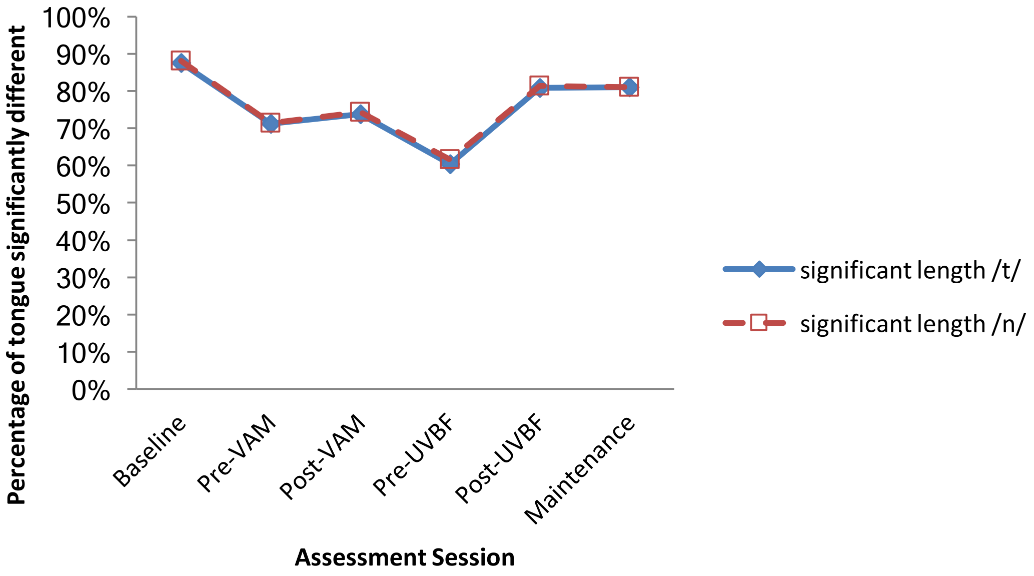


Figure 73 Proportion of the Tongue Identified as Being Significantly Different between /t/ and /n/ in the DEAP

The mean and maximum width between /t/ and /n/ were calculated. As described in the method section, where there is significance reported, the distance value reported refers to the mean radial "width" of the zone centred on the five+ contiguous fanlines for which significance was found, flanked by five associated crossovers. Where there is no significant zone, the mean width for all fanlines is reported (see Figure 74). The largest mean width across the significant zone was found post-UVBF (3.6mm), with the smallest mean width also found post-VAM (2.6mm).

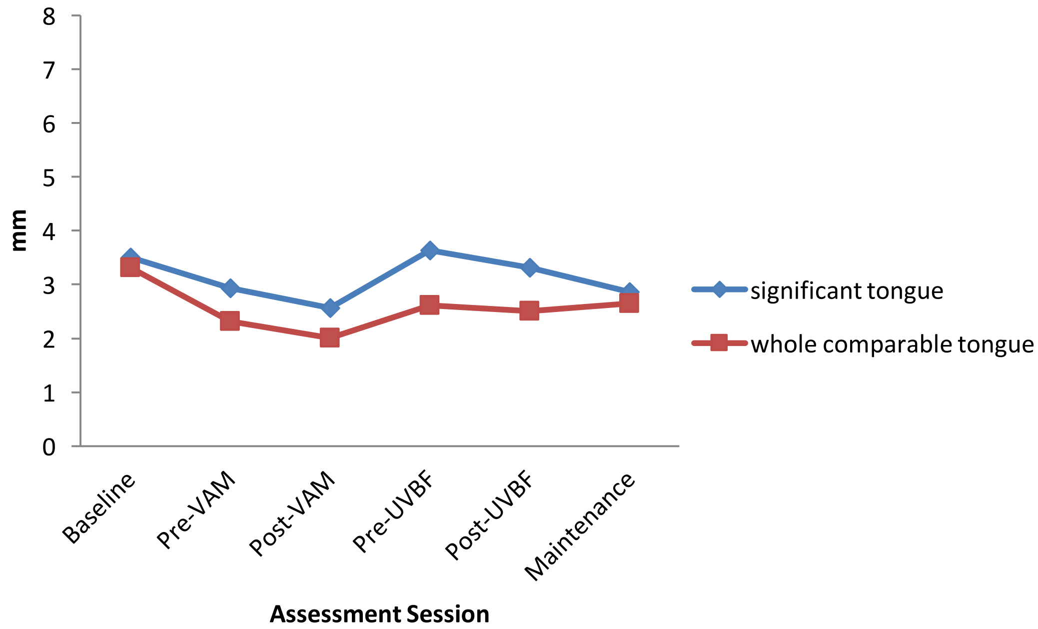


Figure 74 Mean width between /t/ and /n/ across the significant tongue and comparable tongue lengths

The maximum width is reported in Figure 75 for the whole comparable tongue. The largest maximum difference across the comparable tongue was found in baseline (5.3mm), with the smallest maximum difference found post-VAM (4.2mm).

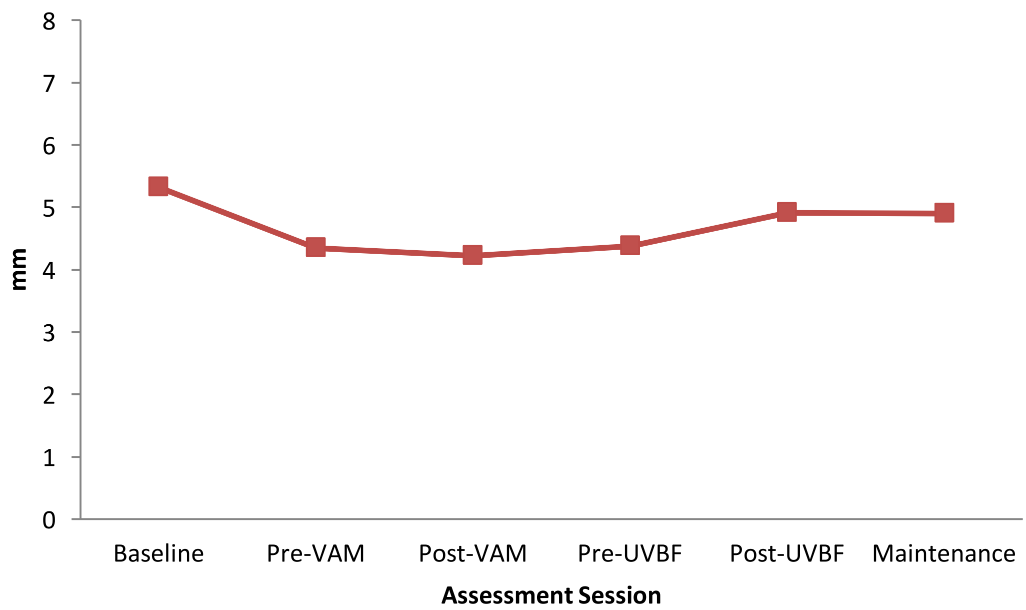


Figure 75 Maximum width between /t/ and /n/ across the whole comparable tongue length

4.2.1.6 Quantitative analysis of /n/ vs. /k/

Figure 72 shows the total and comparable tongue lengths for /n/ and /k/ from the DEAP. The longest visible tongue length for /k/ was found in pre-VAM (58mm), with the shortest visible tongue for /k/ found in the pre-UVBF session (38mm total length, 35mm comparable length). The tongue length for age-matched peers for /k/ is between 62 and 66mm, thus indicating that there is less visible tongue in Andrew's data for /k/ than in the images of his peers. For /n/, the longest visible tongue was found in the pre-VAM session (whole tongue 61mm, comparable tongue 58.3mm) and the shortest was also in pre-UVBF (35mm). When matched against the average length for /t/ in typically developing peers, this is also below the mean tongue length of 62 to 65mm.

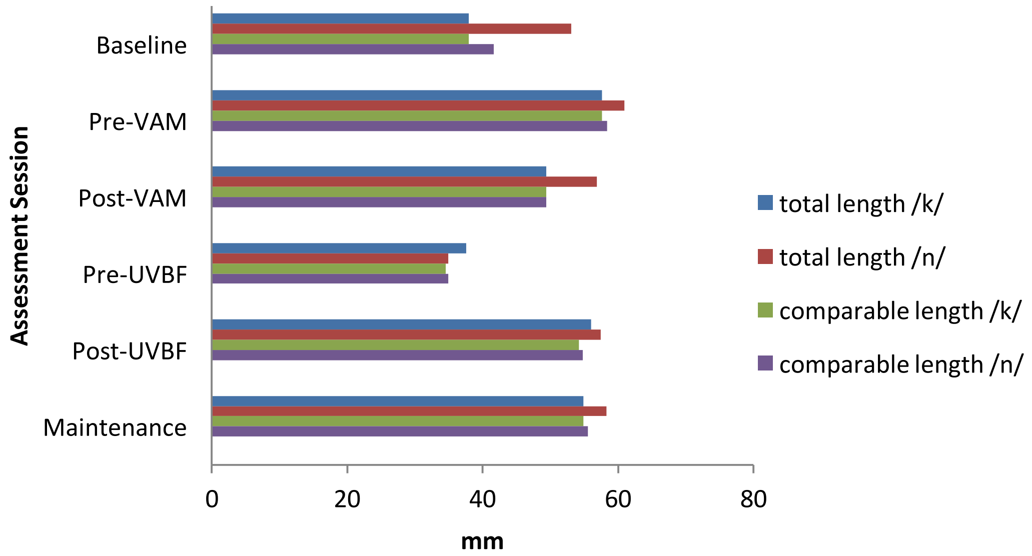


Figure 76 Andrew's Tongue Length Visible for /n/ and /k/ in the DEAP phonology subtest

The proportion of the tongue that is significantly different was calculated using the paired t-test function in AAA. As /n/ was phonetically transcribed as being velar, it was expected that there would be no difference between tongue shapes for /k/ and /n/. As expected, there was no statistically significant difference between /k/ and /n/, indicating that productions of /n/ and /k/ are both produced in the same place of articulation (which, given the production of /k/, can be safely assumed to be velar).

The mean width between /k/ and /n/ were calculated. As there was no significant difference between /k/ and /n/ in any of the six sessions, only the maximum and mean widths for the whole comparable tongue are presented, as descriptive measures. In Figure 77, the mean width is below 2mm in all six sessions, with the smallest difference found in post-VAM (0.6mm) and the largest difference found in the maintenance session (1.5mm). As there is no difference across any of the sessions, this suggests that the place of articulation for /n/ is the same from baseline to maintenance. This lack of evidence of improvement or of any contrast between the two sounds differs from the evidence from the PTCC scores and the perceptual evaluation results, where listeners detected improvement in production of /n/ overall from baseline to maintenance.

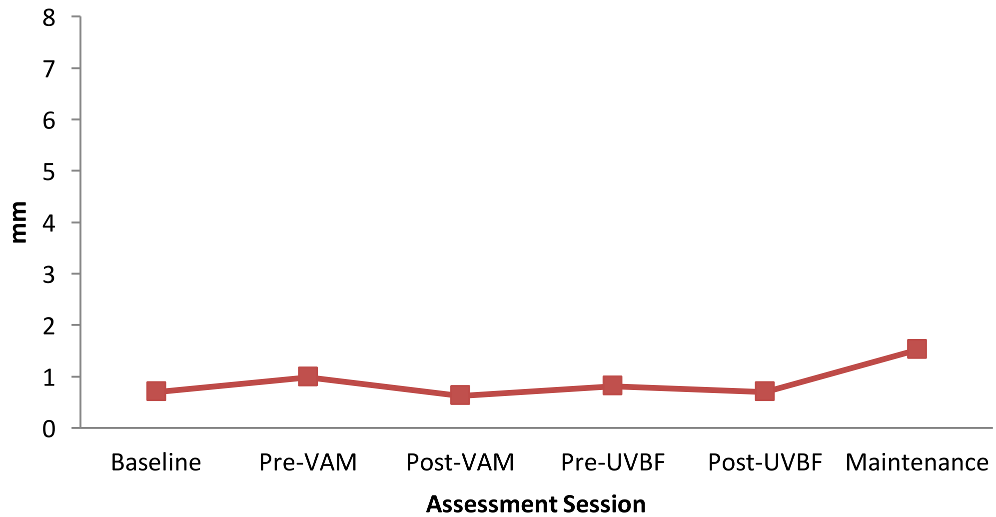


Figure 77 Mean width between /n/ and /k/ across the comparable length of the tongue in the DEAP data

Figure 78 shows the maximum width found on any of the radii in the fan grid across the whole comparable tongue in each session. The largest maximum distance was found in maintenance (4.2mm), with the lowest found pre-UVBF (1.6mm). The maintenance session might therefore have some weak articulatory indications that /n/ and /k/ are beginning to differentiate, but it cannot be ruled out that this is just random noise, and no significant difference in articulation will be reported for this session from this dataset.

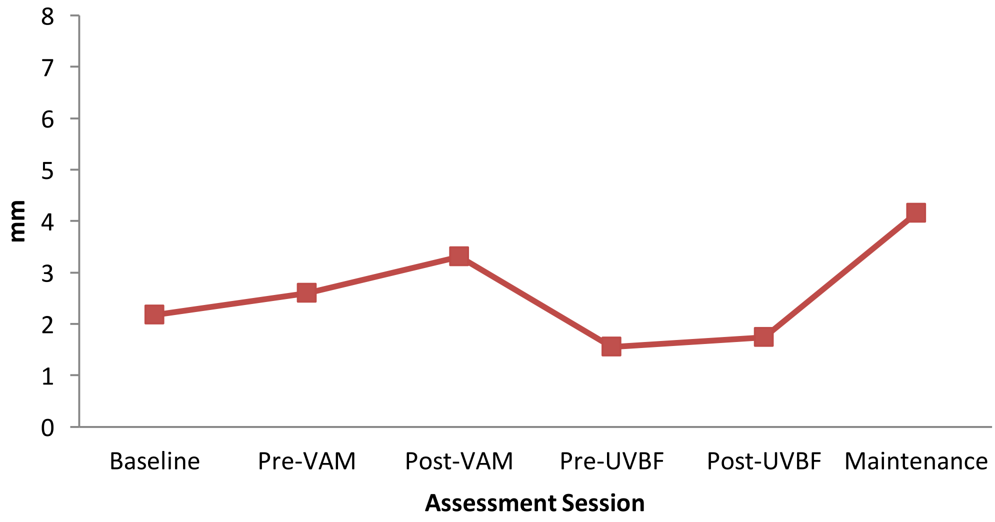


Figure 78 Maximum width between /n/ and /k/ across the whole comparable tongue

4.2.1.7 Andrew: summary

Qualitative and quantitative analysis of the ultrasound data identified covert errors in Andrew's productions of alveolar and velar nasal stops. Covert contrasts ("covert" given the expectation of complete merger based on clinical notes and transcription) were identified in the production of /n/ when compared to /ŋ/ as measured in the minimal pairs data, with statistical analysis showing that the large significant zone of difference visible in the tongue images in Figure 58 was mainly in the tongue root and dorsal regions. Significant differences (supporting the transcription of /n/ as an incorrect velar) were found between the production of /t/ and /n/ in the DEAP, with, additionally, no difference in tongue shapes for /k/ and /n/ across all six assessment sessions.

The articulatory errors identified are in line with some of the errors identified for this client group using EPG, such as retraction, overuse of the tongue dorsum and double articulations. Whilst a significant zone was identified in each of the six sessions, the articulatory size of the contrast was small. When compared to the typical data presented in section 4.1.2.1 (assuming that a contrast between alveolar and velar nasals acts in the same way as a contrast between alveolar and velar plosives), it is evident that the mean and maximum widths are substantially lower than the typical norms.

The tongue length data showed that most of the sessions showed less visible tongue length than expected from age-matched peers' data of typically developing children, indicating poor image quality mostly likely due to tongue tip data missing from the images. This was indeed apparent on raw images with a larger mandible shadow than is normally obtained. One possible reason for this is Andrew's small submental surface due to a small jaw in which the chin is closer to the pharynx than normal. It was also suspected that the probe was not always in midsagittal position. Due to facial asymmetry as a result of his hemifacial microsomia and microtia, it was particularly difficult to fit the headset in a straight position and therefore the probe was likely to be off centre some of the time. It is also evident from a face-on camera view that there is deviation of the jaw to the left when speaking and therefore there is apparent probe movement.

4.2.2 Craig

Craig was being treated for problematic placement of velar plosives, therefore this section will concentrate on the tongue shapes for /k/ and /g/. Firstly, it will report the qualitative and quantitative analyses of velar plosives to velar nasals and also make a comparison against the voiceless vs. voiced velar plosives from the untreated wordlist. This will explore the similarities and differences in velars when incorrectly produced (the oral plosives), as against the velar nasal, which was transcribed correctly. In the baseline session, productions of /k/ were transcribed as glottal stops and /g/ transcribed as either [d] or [n]. The untreated wordlist only included the velar targets. Secondly, it will report a qualitative and quantitative analysis of alveolars vs. the fronted tokens of the velar plosives, using the DEAP phonology subtest (Dodd et al. 2002). This DEAP comparison was made due to the lack of better, more targeted minimal pair data for Craig, and because Craig's productions of velar stops were inconsistently transcribed as alveolar.

4.2.2.1 Analysis of velar plosives and velar nasal stops: untreated wordlist

Figure 79 shows the average tongue shapes for /k/ (blue), /ɡ/ (green) and /ŋ/ (pink) for each of the six assessment sessions. In typical speech, it would be expected that all three tongue shapes would be near-identical, if velars were to act in the same way as alveolar oral and nasal stops. Craig's production of /ŋ/ was correct throughout all six assessments, which acted as a facilitative environment for therapy in order to elicit the velar plosives. Assuming that velar nasals and plosives should have identical tongue shapes, the shape for /ŋ/ can act as a benchmark for comparison with velar plosives.

In the baseline and pre-VAM sessions, tongue shapes for /k/ and /ɡ/ are the same with the tongue tip raised and a clear difference in shape from the velar nasal which has the tongue dorsum raised and the tongue tip lowered. In other words, the obstruents are wrong in similar ways to each other and the tongue shapes support the transcription of these targets as being fronted velars. By the post-VAM session, the obstruents have improved in their production so that all three shapes are near identical (velar), with similar patterns in the remaining sessions. This confirms perceptual analysis that showed an increase in PTCC post-VAM and listeners in the perceptual evaluation selected post-VAM as being closer to target than pre-VAM. In the post-UVBF session, there is apparently a large variation in the production of velar nasals, indicated by the large standard deviation.

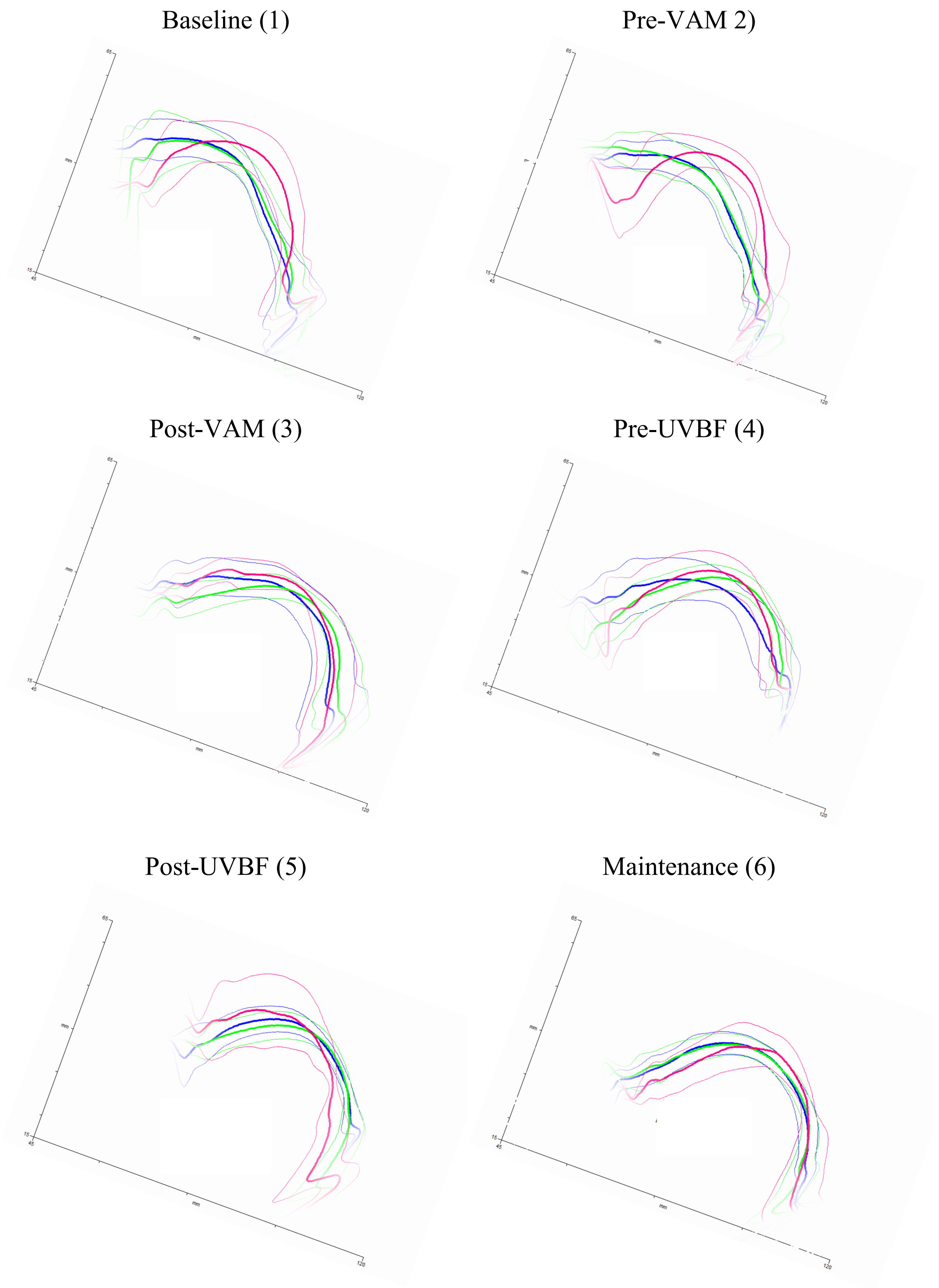


Figure 79 Untreated velar wordlist: comparison of averaged /k/ (blue) /g/(green) /ŋ/(pink)

When looking at the data for /ŋ/ before averaging (Figure 80), showing a huge amount of variation, it is obvious however that there is a large amount of headset movement rather than variation in speech production.

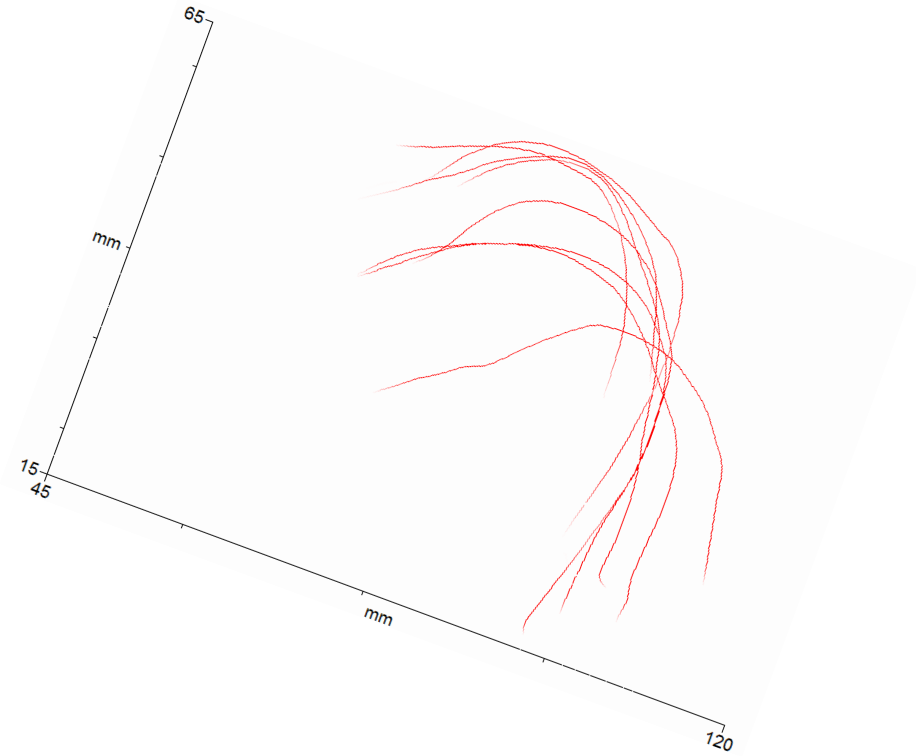


Figure 80 Apparent variation in velar nasals post-UVBF due to headset-probe instability

4.2.2.2 Quantitative Analysis of /k/ vs. /ŋ/

Figure 81 shows the total and comparable tongue lengths for /k/ and /ŋ/ from the untreated velar wordlist. With the length in typically developing children's tongues ranging from 55mm at age six years to 57mm at age seven years, it is evident from Figure 81 that Craig's visible tongue length is considerably shorter than expected for his age across the course of the nine months of recordings (from baseline to maintenance). The longest visible tongue for both consonants was found in the maintenance sessions (/k/ 45mm total length, 42mm comparable length; /ŋ/ 44mm total tongue, 42mm comparable tongue). The shortest visible tongue was found in post-UVBF (/k/ 29mm comparable length, /ŋ/ 29mm comparable length). Within this comparison there is at least 10mm less visible tongue than his age-matched peers.

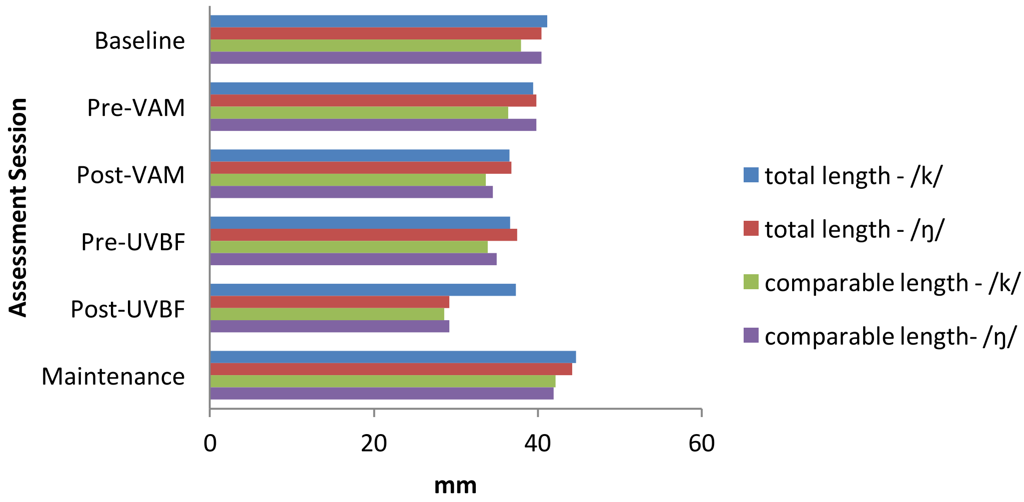


Figure 81 Craig's Tongue Length Visible for /k/ and /ŋ/ in the Untreated Velar Wordlist

It is evident from Figure 82 that there was a significant zone of at least 47% at baseline and 56% at pre-VAM. In these sessions, there is a significant difference between the oral and nasal velar stops. However, following treatment using VAM, from post-VAM through to maintenance there is no difference between /k/ and /ŋ/.,

These comparisons suggest that /k/ is incorrect pre-VAM, as was expected from the clinical notes and baseline transcription, but is produced in an appropriate velar placement (assuming that the velar nasal is correct) post-VAM and thereafter. This change confirms phonetic transcriptions, with improvement in PTCC post-VAM. However, PTCC never reached 100%, suggesting that /k/ did not sound perceptually correct in all sessions, and it should be remembered that even though t-tests have shown no difference between /k/ and /ŋ/, this does not mean there are identical in tongue placement, even before other articulatory factors are taken into account. Around 50% of the comparable lengths of the oral and velar stops is significantly different pre-therapy, and no /k/ productions were transcribed as correct.

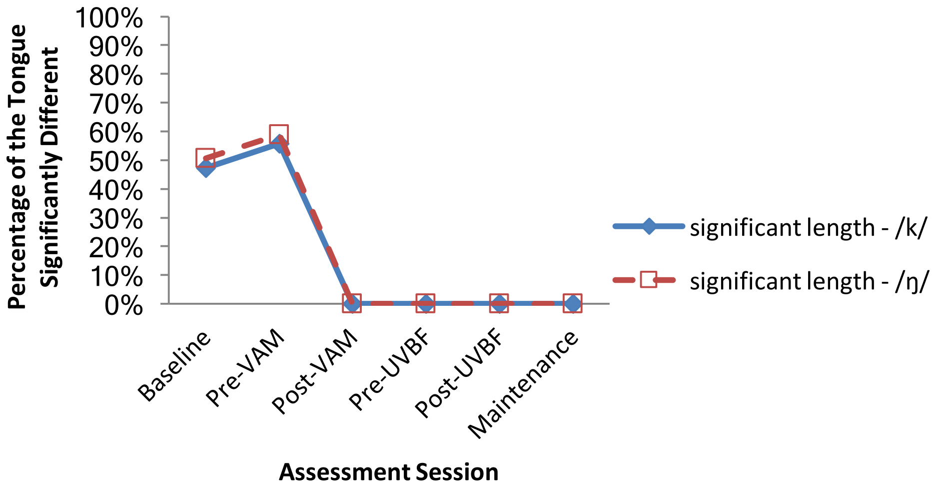


Figure 82 Proportion of the Tongue Identified as Being Significantly Different between /k/ and /ŋ/

As there were only significant differences between /k/ and /ŋ/ within the baseline and pre-VAM sessions, the mean width across the significant tongue can be displayed only for these two sessions in Figure 83 (6.3mm at baseline; 6.8mm pre-VAM). Across the comparable parts of the tongue, however, descriptive figures can be given for all the sessions. The largest mean width is seen in the pre-VAM session (5.1mm) and the smallest in post-VAM (1.3mm).

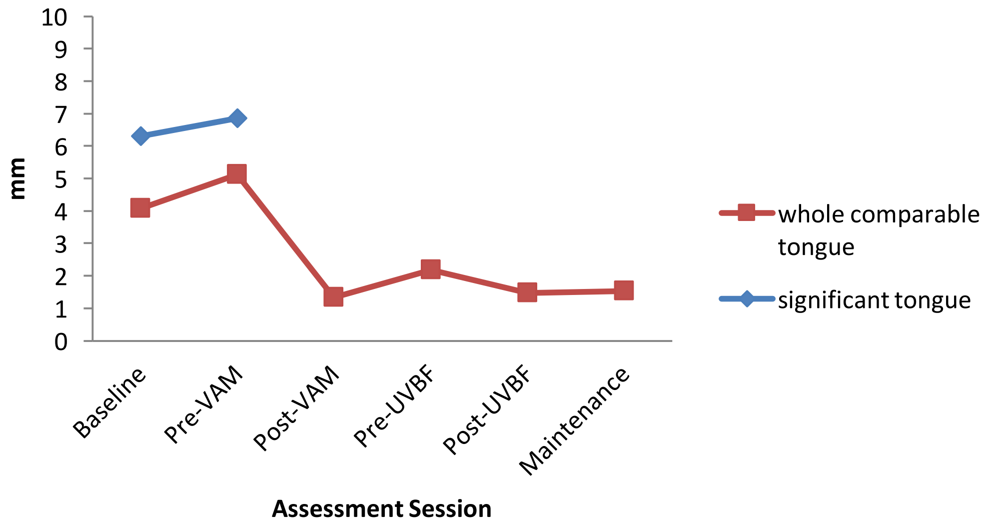


Figure 83 Mean Width between /k/ and /ŋ/ across the whole comparable tongue and the significant zone

The largest maximum width (i.e. for a single radial comparison) across the whole comparable tongue (see Figure 84) is 9.1mm (pre-VAM) and the smallest maximum width is 2.0mm, also found in the post-VAM session. Although there is a small difference between /k/ and /ŋ/ from post-VAM to maintenance (maximum 4mm), these sessions have no significant difference in tongue shape.

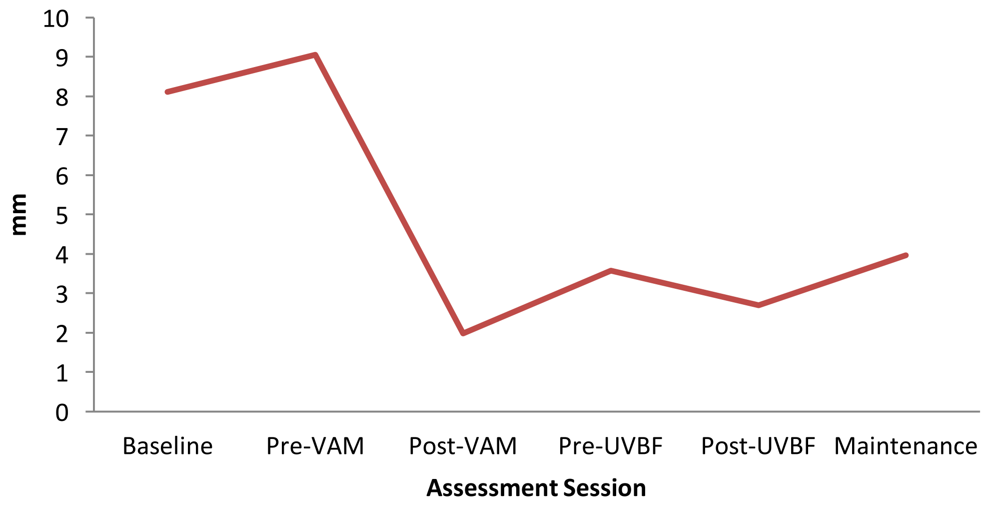


Figure 84 Maximum Width between /k/ and /ŋ/ across the whole comparable tongue

4.2.2.3 Quantitative Analysis of /g/ vs. /ŋ/

As /k/ and /g/ behaved differently during the pre-therapy sessions, it is also useful to look at the comparison of /g/ and /ŋ/. Figure 85 shows the total and comparable tongue lengths /g/ and /ŋ/ from the untreated velar wordlist. Similar to the /k/ vs. /ŋ/ comparison, tongue lengths here are noticeably shorter than expected from the typical norms. As in the /k/, /ŋ/ comparison, the maintenance session had the longest visible tongue for /g/ (42mm) and /ŋ/ (42mm), more than 10mm shorter than 55mm which would be expected for his age. The shortest visible tongue was post-UVBF, indicative of the poor image quality in this session.

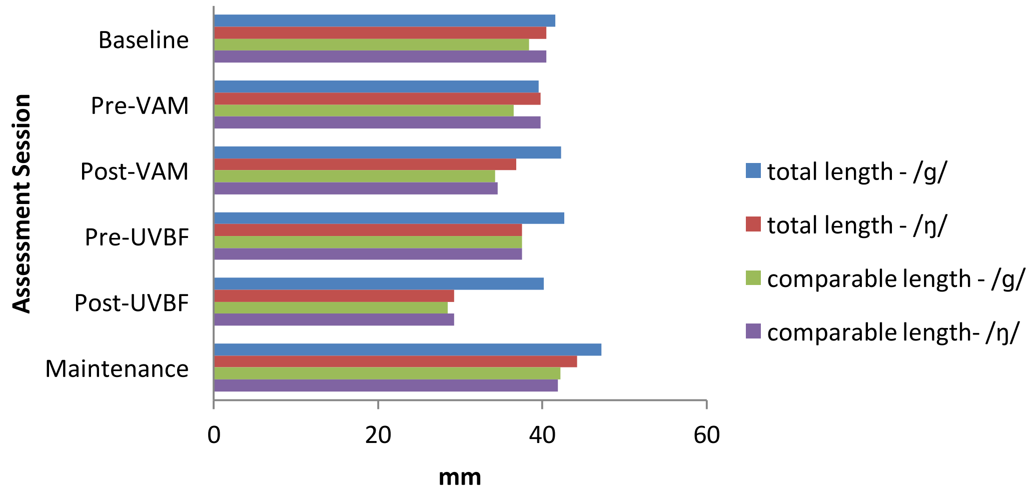


Figure 85 Craig's Tongue Length Visible for /g/ and /ŋ/ in the Untreated Velar Wordlist

Quite differently from the /k/, /ŋ/ comparison, the zone of significance between /g/ and /ŋ/ across the comparable tongue is 100% (see Figure 86), where the zone of significance was only around 50% for /k/ and /ŋ/. As /g/ was more frequently transcribed as alveolar than /k/ was in baseline, this is not surprising: /g/ would be expected to have an alveolar articulation as well as lacking velar constriction. In pre-VAM, this reduces to 56% for /g/, which is the same as /k/. Likewise, post-VAM to maintenance, there is no difference between /g/ and /ŋ/, indicating an improvement in Craig's production of both velar plosives, confirming phonetic transcriptions.

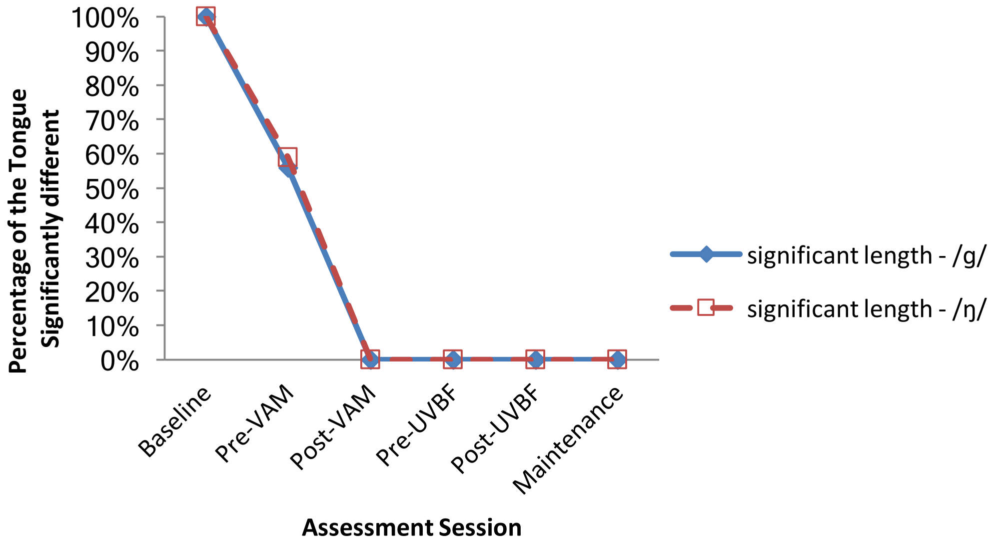


Figure 86 Proportion of the tongue identified as being significantly different between /g/ and /ŋ/

Despite being differentiated over all their length, the mean width between /g/ and /ŋ/ is somewhat shorter in baseline (3.6mm) than it was between /k/ and /ŋ/ (6.3mm), which were different over about half their comparable length. Pre-VAM the mean width of the significant zone is roughly the same. Equally, the mean width post-VAM to maintenance is under 3mm and is not significant (see Figure 87).

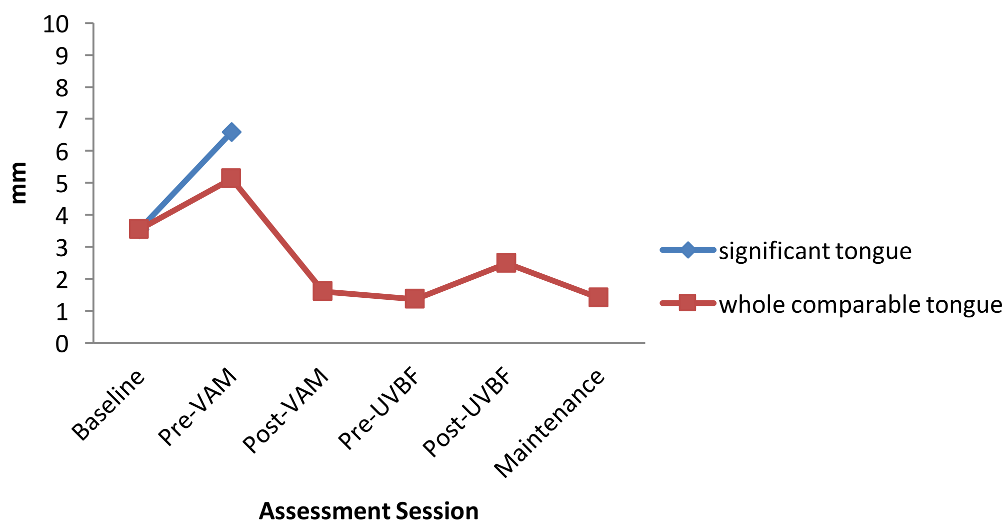


Figure 87 Mean Width between /g/ and /ŋ/ across the whole comparable tongue and the significant zone

The largest maximum width between /g/ and /ŋ/ across the whole comparable tongue (see Figure 88) in pre-VAM (8.5mm), with the smallest width found in pre-UVBF (2.6mm).

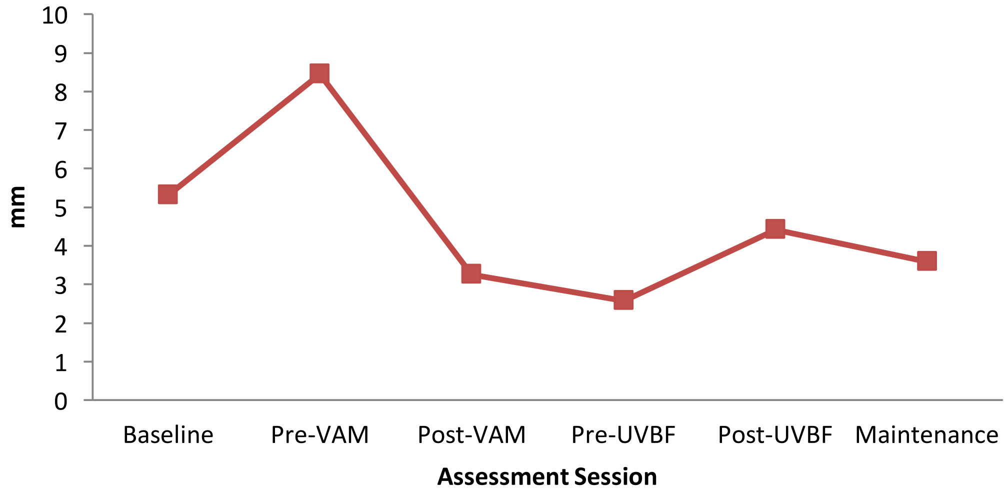


Figure 88 Maximum Width between /g/ and /ŋ/ across the whole comparable tongue

4.2.2.4 Quantitative Analysis of /k/ vs. /g/

Due to differences in transcriptions for /k/ and /g/ during the pre-therapy sessions, and the different results when each of them is compared to /ŋ/, as just shown, a quantitative analysis between /k/ and /g/ was also carried out. Figure 89 shows the total and comparable visible tongue lengths for /k/ and /g/. Whilst the total and comparable lengths are almost identical for baseline and pre-VAM, the total length for /g/ is clearly longer in the remaining sessions. That is, more of the spline for /g/ had over 80% confidence in the automated edge-tracking than for /k/. When comparing all tongue lengths from the untreated wordlist, it is clear that the splined length of /g/ is indeed generally more visible than /k/, with (as seen previously) Craig’s plosives generally being more visible than nasals.

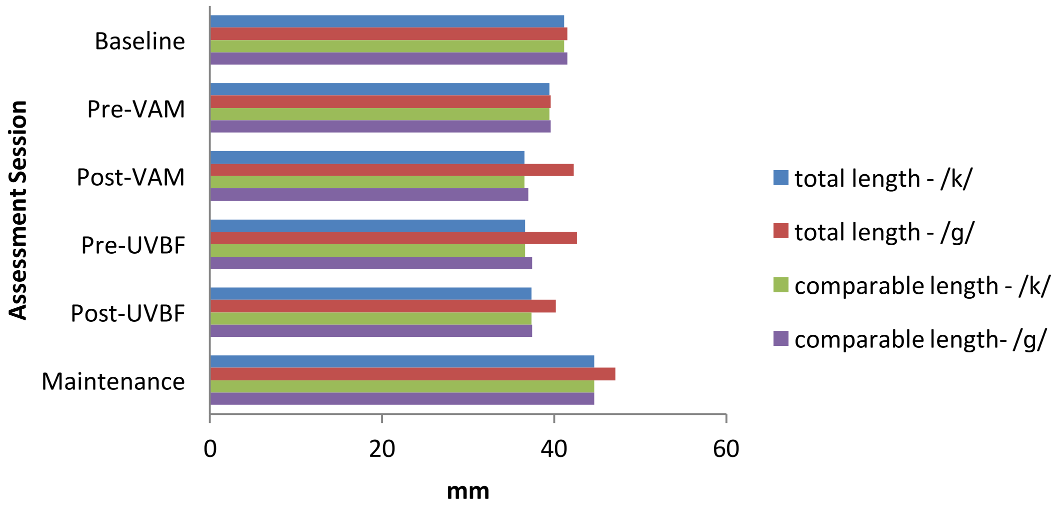


Figure 89 Craig's Tongue Length Visible for /k/ and /g/ in the Untreated Velar Wordlist

Ultrasound analysis provides a surprisingly different finding in the first two sessions, however. In those, whilst /k/ and /g/ were transcribed differently (i.e. /k/ was transcribed as a glottal stop and /g/ was transcribed as [d] or [n]), statistical analysis contradicts these transcriptions with no difference found in the tongue shapes for /k/ and /g/. In fact, both /k/ and /g/ were fronted to alveolar placement, and the apparent backing to glottal (replacement of /k/ by a glottal stop [?]) is revealed as glottal reinforcement on alveolar productions. Post-VAM, there is still no significant difference between /k/ and /g/, and the place of articulation of both has changed from alveolar to velar. However, this change is not without one additional twist. Interestingly there is a 45% (/k/) and 48% (/g/) difference pre-UVBF (see Figure 90) indicating, for the first time, a difference in tongue shape between these two plosives, and just in this one session.

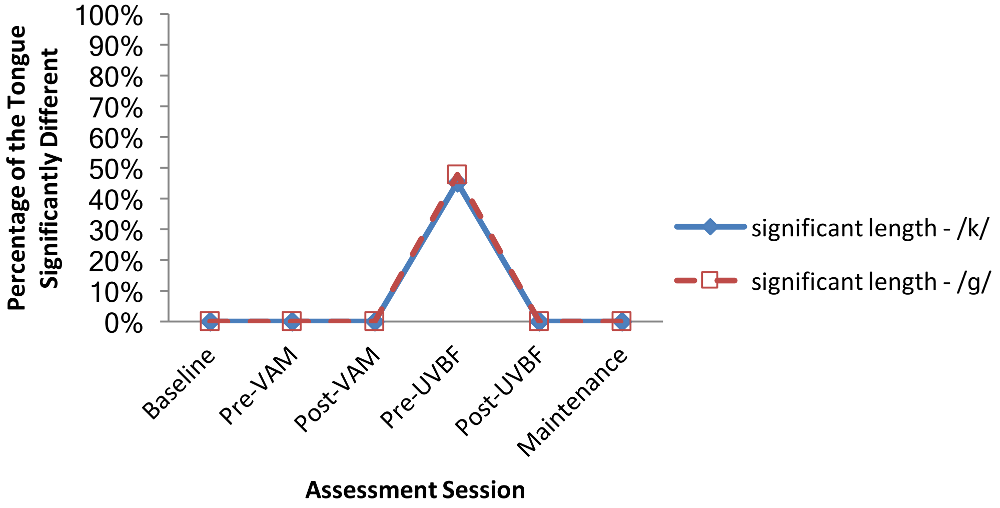


Figure 90 Proportion of the tongue identified as being significantly different between /k/ and /g/

As a significant zone was only found in pre-UVBF, this is the only significant mean width provided in Figure 91 (3.7mm). Whilst this is a significant difference, it is somewhat smaller than the differences reported between /k/ and /ŋ/ and /g/ and /ŋ/.

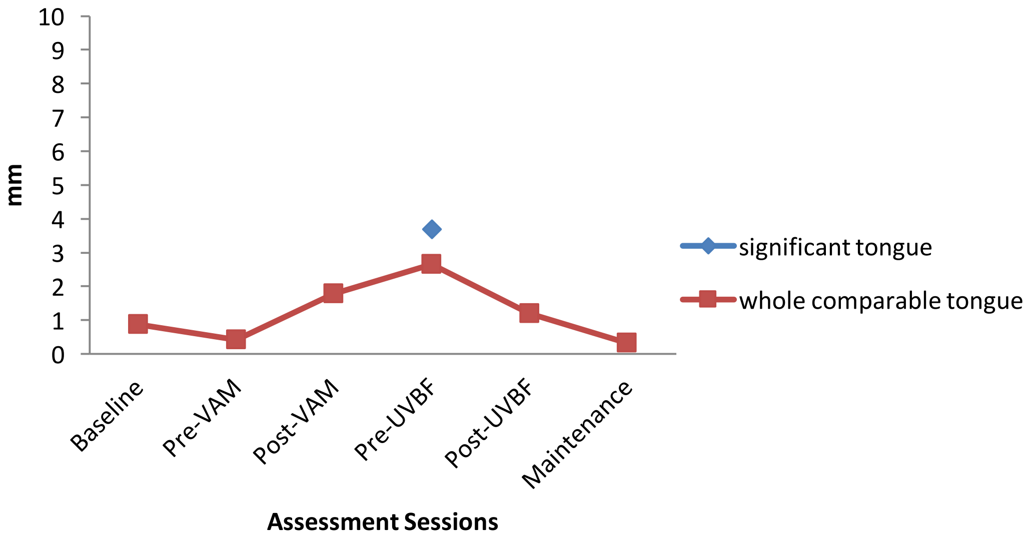


Figure 91 Mean Width between /k/ and /g/ across the whole comparable tongue and the significant zone

Similarly, the maximum width found in pre-UVBF between /k/ and /g/ (see Figure 92) is smaller than in the other two comparisons, indicating more similarity at a descriptive level between voiced and voiceless velar plosives than between velar plosives and nasals, bearing in mind that there is no significant difference identified. It is tempting to see the pre-UVBF difference as the peak in a trend from Pre-VAM to Maintenance, but there is no guarantee that this is anything other than noise in the data.

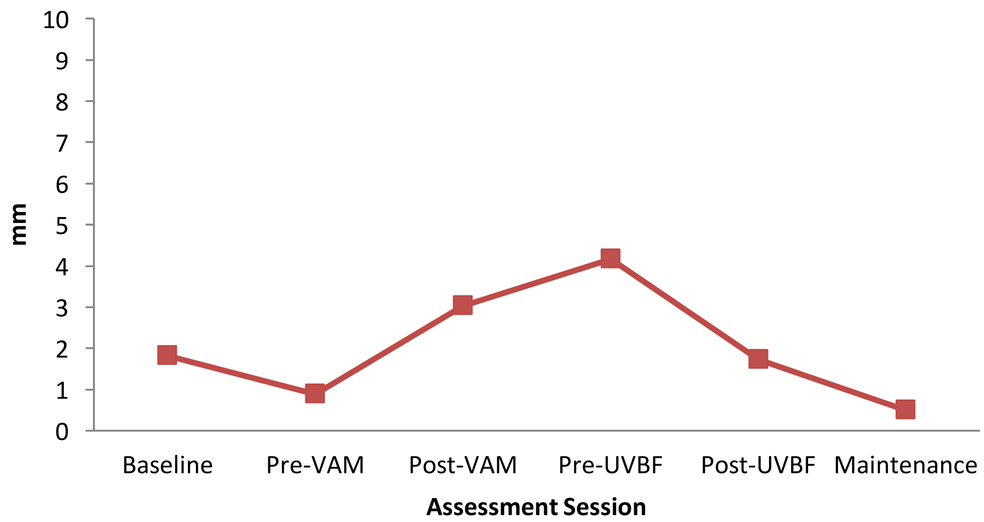


Figure 92 Maximum Width between /g/ and /ŋ/ across the whole comparable tongue

4.2.2.5 DEAP: alveolar and velar productions

As /t/ was also treated in therapy block two (UVBF) as an additional target, and because there were no minimal pairs in Craig's data, a comparison was made between alveolar and velar plosives using the DEAP data. Assuming on the basis of what we saw above that voicing is not a factor, this allowed for a qualitative and quantitative analysis between /t/ and /d/ averaged (red) and /k/ and /g/ averaged (blue) (see Figure 93). A difference in tongue shapes between alveolar and velar placement would be expected on the basis of the original transcriptions, however having seen the covert alveolar placement for /k/ as well as the transcribed fronting for /g/, we would now expect no difference if /t/, likewise, was produced as a glottally-reinforced alveolar rather than a pure glottal stop. Pre-therapy in both the baseline and pre-VAM sessions there is indeed no statistical difference between the tongue shapes. This is interesting, as now we can see that both voiceless alveolar and velar targets were often transcribed as glottal stops in the phonetic transcriptions, yet have (sometimes correct) covert lingual movements in the ultrasound data. This, along with phonetic transcriptions suggests that Craig was velar fronting pre-therapy. This is not a typical cleft-type characteristic. However, it may be the case that generally, as here with Craig, velar fronting errors are often not perceived correctly due to glottal reinforcement on voiceless alveolar stops, something that is indeed typical of the speech of individuals with CP. From post-VAM onward, there is a contrast between alveolar and velar, with the biggest difference seen in post-UVBF.

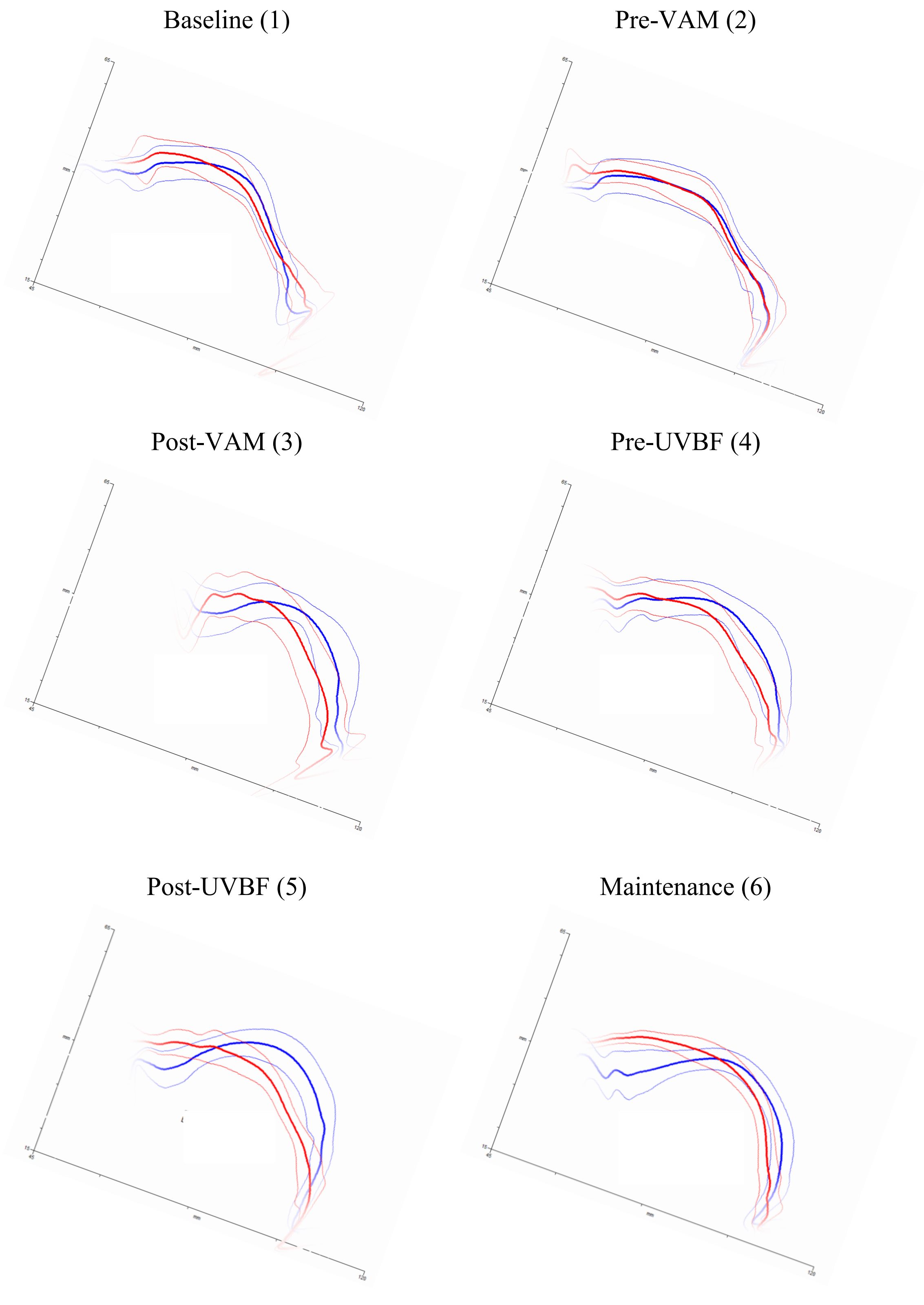


Figure 93 Average alveolar (red) and velar plosives (blue) in the DEAP across all six assessment sessions

Similar to the velars in the untreated wordlist, and even when alveolar consonants are included, visible tongue lengths are somewhat shorter than expected for Craig's age (see Figure 94). Based on the typical norms derived from the ULTRAX data, it would be expected that Craig's visible tongue length for alveolars should be between 53mm (norm for age six years) and 56mm (age seven norms), and for velars should be between 55mm (aged six) and 57mm (aged seven). This would either suggest that either Craig's tongue is generally shorter than his peers or that the image quality is poorer than that of typically developing children, with the latter more likely.

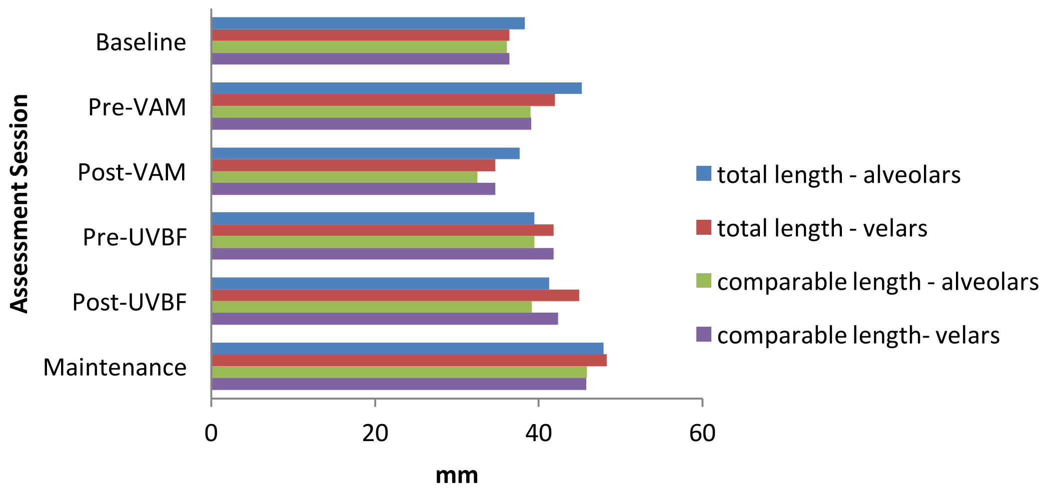


Figure 94 Craig's Tongue Length Visible for alveolar and velar plosives in the DEAP phonology subtest

As Craig's production of velars and alveolars improve during the intervention, a larger contrast between the two would be expected in later sessions. Figure 95 shows no significant difference found between alveolars and velars pre-therapy, indicating a merger in lingual patterns. A large increase from zero to 60% is found after therapy block one, using Speech Trainer 3D, and there is stability between the two therapy blocks (a five-week period of no intervention). A further large improvement, as indicated by the % increase, is found post-UVBF where the velar vs. alveolar difference is evidenced for the whole tongue visible in the images, quantified as a significant zone of 100%, which remains stable three months post therapy in the maintenance session.

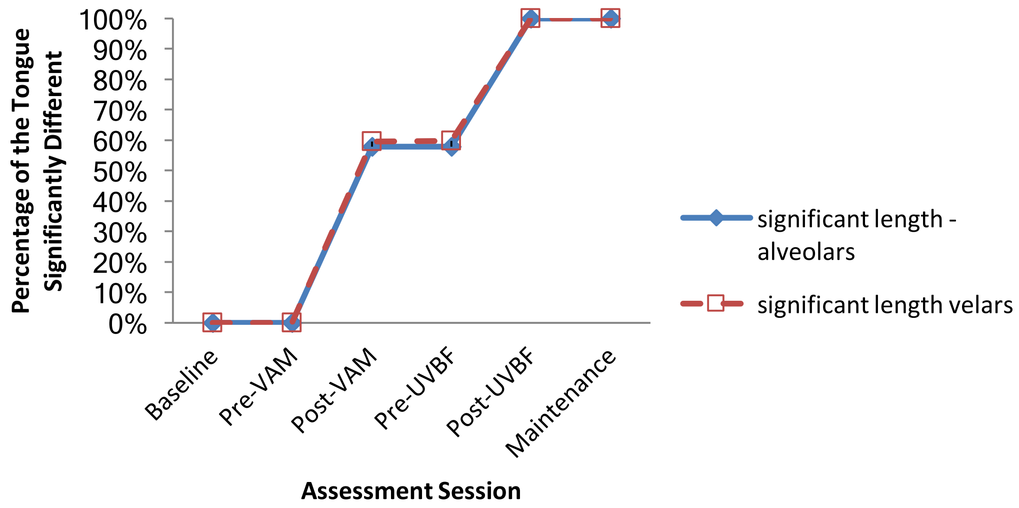


Figure 95 Proportion of the tongue significantly different between alveolars and velars in the DEAP

Figure 96 shows the mean width between alveolar and velar plosives across the whole comparable tongue for all six sessions and the significant tongue from post-VAM through to maintenance, since there was no significant difference found at baseline and pre-VAM.

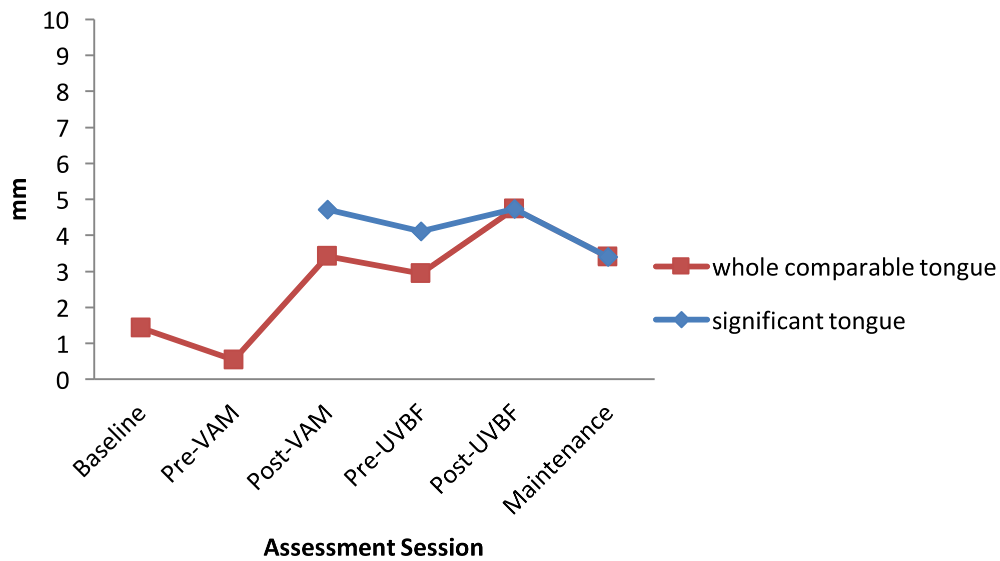


Figure 96 Mean Width between alveolars and velars in the DEAP, across the whole comparable tongue and the significant zone

As the length of the significant zone increases, the maximum width between the two targets also increases (see Figure 97). The largest maximum width is found in the maintenance session with 7.8mm difference between alveolar and velar plosives. The mean, however is only 3.4mm within the maintenance session, with the largest mean found in the post-UVBF session (4.7mm).

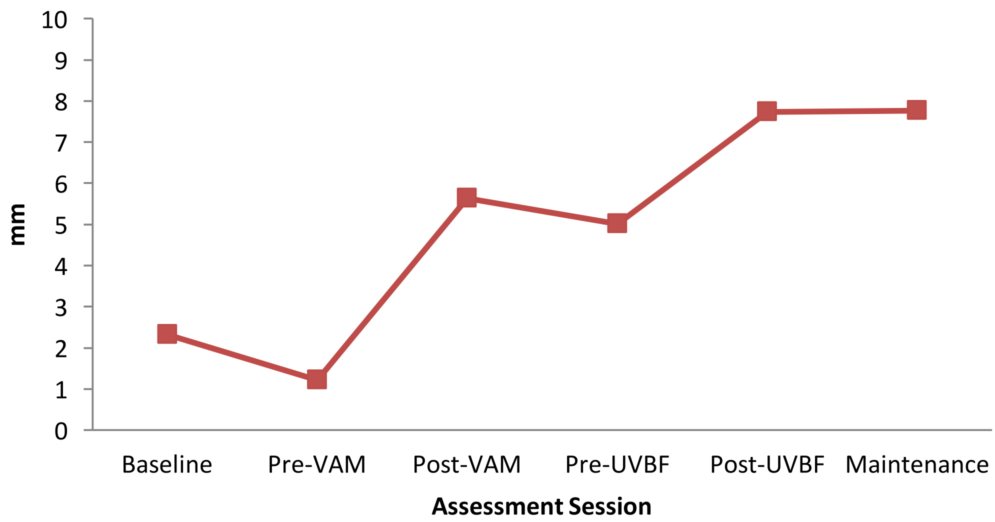


Figure 97 Maximum Width between alveolars and velars in the DEAP, across the whole comparable tongue

4.2.2.6 Craig: summary

Although Craig's data was of lesser quality to his TD peers, qualitative and quantitative analysis was still possible. These measures identified covert articulations in Craig's speech. More specifically, when alveolars or velars were identified as a glottal stop in phonetic transcription, alveolar lingual movements were identified by ultrasound. Two errors are evident. Firstly, Craig is velar fronting velar plosives to alveolar placement. Secondly, there is glottal reinforcement on voiceless alveolar plosives (inappropriate for a syllable onset) and the alveolar lingual movements are not perceived through phonetic transcription. The PTCC and PCC scores increase as treatment progressed, both in the untreated wordlist and the DEAP, and the ultrasound data confirms that the contrast between alveolar and velar plosives became significant and the significant width increased in line with the transcription data.

Analysis of the velars in the untreated wordlist show that pre-therapy there are clear differences between velar plosives and nasals. Significant differences between /k/ and /g/ were also surprisingly found in the pre-UVBF session with a mean width of 3.7mm.

4.3 Summary of Articulatory Analysis

This section has provided an articulatory analysis for Andrew and Craig's data, with a comparison to the data of age-matched TD peers and to phonetic transcriptions. Due to structural abnormalities in Craig and Andrew, for example small jaw and facial asymmetry, the data for both children was of lesser quality to their TD peers. However, qualitative and quantitative measures were still possible, which supplemented phonetic transcriptions by providing additional information including covert error and covert contrasts. Quantitative and qualitative measures allowed for a comparison of pre- and post-therapy tokens to allow for longitudinal measures of change.

The following chapter will present a discussion of the findings from this chapter and previous chapters to address the theoretical and clinical implications for the use of UVBF and VAMs for children with CP.

5 Discussion

The previous chapters in this thesis have presented the results from three methods of analysis, to measure the effectiveness of VAMs and UVBF therapy for two children with repaired SMCP. The research questions in the treatment study related to an increase in PTCC scores to measure improvement in speech outcomes post-therapy. The perceptual evaluation aimed to measure multi-listener judgements, asking whether listeners would select more post-therapy tokens as 'closer to the English target' than pre-therapy tokens in six different comparisons. Finally, the research questions in the articulatory analysis chapter related to the diagnostic use of UTI, whether it was able to identify any covert errors and whether it was able to detect any longitudinal change in ultrasound data.

The current section will firstly discuss the therapy outcome measurements with regards to all methods of analysis presented in the previous chapters, i.e. phonetic transcription, multi-listener perceptual evaluation and instrumental analysis, as assessment was carried out prior to therapy using UVBF. Secondly, a discussion of the therapy design and a comparison of visual articulatory models and ultrasound visual biofeedback will be presented with future implications for assessment and therapy using these tools for individuals with cleft palate.

5.1 Summary of the Key Findings

Previous studies (Gibbon 2004; Howard 2004) have identified covert lingual patterns using EPG, with the RCSLT (2005) recommending EPG as a clinical tool for the remediation of SSDs associated with CP. However, Lee et al. (2009) highlight the lack of robust procedures to measure the effectiveness of EPG therapy with only one RCT found in their literature search. Lee et al. (2007) highlight some of the problems with using EPG, such as the need for stable dentition and requirement of periods with no surgical treatment planned. Similarly, UTI has been used to investigate the compensatory articulations in speakers with CP (Gibbon and Wolters 2005; Bressmann et al. 2011), however its therapeutic application has not previously been tested. In comparison to VBF techniques, off-line VAMs, commercially available on iPads, such as the Speech Trainer 3D app (Smarty Ears 2011), provide a context for lingual movement in relation to passive articulators. However, Speech Trainer 3D does not provide real-time biofeedback like UTI and its application for treating SSDs remains to be tested. This thesis aimed to investigate the diagnostic use of ultrasound and the clinical effectiveness of VAMs and UVBF for individuals with CP.

Using a single-subject multiple-baseline design with an ABACA treatments design, two children, Andrew and Craig, received two blocks of motor-based therapy. Firstly, with the app Speech Trainer 3D (Smarty Ears 2011) used as an off-line Visual Articulatory Model (VAM), to provide context for learning new articulatory gestures, and secondly augmented with ultrasound visual biofeedback (UVBF). Each block consisted of eight one-hour long therapy sessions and six assessment sessions, in which speech measures including the DEAP Phonology subtest (Dodd et al. 2002), target-specific wordlists and the ICS (McLeod et al. 2012) were repeated. For Andrew, the therapy target was /n/ and for Craig the therapy targets were velar plosives, with the addition of /t/ in therapy block two, to reinforce a /t/, /k/ contrast. There were three methods of evaluating the effectiveness of therapy, summarised below.

Firstly, perceptual assessment of PTCC scores and PCC scores were derived from phonetic transcriptions of the target specific wordlists and DEAP Phonology subtest. The ICS and feedback questionnaires were used to measure intelligibility and the

GOS.SP.ASS'98 (Sell et al. 1999) was used as an additional speech outcome, which was completed by the CLP specialist SLT two (Craig) or four months (Andrew) before commencing on the study and nine months after the project ended. Research questions referred to whether PTCC scores, PCC scores and intelligibility would increase post-therapy. Results showed an increase in PTCC from the target specific untreated wordlists for both children from baseline to maintenance, with a larger increase in PTCC scores post-VAM than expected. In fact, more improvement was found in the first therapy block with VAMs than in the second therapy block with ultrasound, contrary to the hypothesis which expected stability in therapy block one due to little evidence to support the advantage of using VAMS for learning new articulations. For Andrew, the majority of his errors with /n/ were retraction to velar placement, with some double articulations transcribed. For Craig, velar productions were much more variable in baseline and pre-VAM sessions, in both the untreated wordlist and the DEAP, including velar fronting, glottal replacement and glottal reinforcement. Intra-rater reliability for Craig showed very good agreement based on statistical analysis, apart from the post-UVBF session which had moderate agreement. An average of 96% of the data was reliably in agreement, in line with Shriberg and Lof (1991)'s agreement levels. However, for Andrew listener agreement was variable and did not adhere to agreement levels. Similarly, inter-rater reliability was below the levels suggested by Shriberg and Lof for both Andrew and Craig, with intermediate to good agreement suggested by Kappa scores. This highlights difficulties with listener agreement in complex speech disorders such as those in CP speakers. The GOS.SP.ASS'98 showed generalisation and maintenance of all lingual targets for both children, indicating that they had learned their therapy targets. When using a motor-based approach, it is expected that true learning will take place with the use of Knowledge of Results (KR) and Knowledge of Performance (KP) feedback, leading to not only the acquisition of a target sound but also the retention and generalisation into untreated targets (Maas et al. 2008; Preston et al. 2014). As this study adopted a motor-based approach for therapy, it was, therefore, expected that both Andrew and Craig would continue to generalise their targets after therapy ceased.

For Andrew, there was no indication of change in intelligibility from the ICS, however for Craig, change was found in intelligibility to all listener types. Despite a greater rate of change with Speech Trainer 3D, in the three-month post-therapy questionnaire, both children reported that they preferred ultrasound to Speech Trainer 3D because they were able to see their own tongues moving in real-time.

Secondly, to overcome some of the difficulties with intra- and inter-rater reliability, a perceptual evaluation of multiple phonetically-trained listeners using a modified two-alternative forced choice design was employed. This allowed for a wider view in terms of speech improvement by using listeners who were not expert in CP, with previous literature suggesting that naïve listeners offer real-life significance to assessments (Sell 2005) and provide a useful adjunct to specialist SLT assessment for acceptability measures (Bagnall and David 1988). Listeners were asked to select the token that was closer to the target from a choice of two tokens (earlier/later in time). Research questions asked whether listeners would select the later time-point (i.e. post-therapy) as being closer to the target than earlier time-points (i.e. pre-therapy) and whether listeners were more confident and if they responded quicker when they selected post-therapy tokens. Results showed that when PTCC scores increased, the listeners in the perceptual evaluation selected post-therapy tokens as being closer to target. When PTCC scores decreased (as in Andrew post-UVBF), the listeners favoured pre-therapy targets, although only one listener showed significant results in this comparison. Previous studies (e.g. Cleland et al. 2015c and Cleland et al. 2017b) found significant improvements in PTCC scores for the majority of speakers, but significant improvement was not found in all speakers. For both children, there was mostly no difference in whether listeners selected pre- or post-therapy tokens in terms of how confident listeners were in their responses or how quickly they reacted, apart from post-UVBF for Andrew which showed a significant difference in confidence when selecting pre-therapy. When listeners selected pre-therapy in the VAM comparison for Andrew, there was a weak correlation between reaction time and confidence. Similarly, for Craig a weak correlation between reaction time and confidence was found when listeners selected post-therapy tokens in the UVBF comparison and when they selected pre- or post-therapy in the BL-M comparison showing a trend of quicker reaction times when listeners were more confident. For the other comparisons, there was no correlation.

Thirdly, and finally, an articulatory analysis of the ultrasound data was carried out using qualitative and quantitative measures, including tongue length and width, to compare two tongue curves, to discover whether ultrasound confirms phonetic transcriptions and whether it provides any additional information to the two previous perceptual methods. Research questions asked whether there were any quantitative differences in tongue shapes from pre- and post-therapy sessions indicating improvement in therapy targets and whether there were any covert errors. Previous studies using EPG (Gibbon et al. 2004; Howard 2004) and ultrasound (Gibbon and Wolters 2005; Bressmann et al. 2011) have identified covert errors, such as retraction, palatalisation and double articulations through instrumental analysis, which were not identified through perceptual measures. The quantitative measures used in the current study were compared to data of age-matched typically developing peers to investigate the anatomical differences in submental space between typically developing (TD) children and the two speakers presented in this thesis with submucous cleft palate (SMCP). By measuring the length of the visible tongue, this also allowed for a comparison of image quality between TD children and those with SMCP. Ultrasound analysis supplemented phonetic transcriptions, with the additional information of covert errors, such as a covert lingual merger of /t/ and /k/ (transcribed in perceptual assessment as glottal replacement, but with the addition of ultrasound more accurately transcribed as alveolar and velar voiceless plosives with glottal reinforcement), and covert contrast (for Andrew's production of /n/ /ŋ/ minimal pairs). Image quality was poorer for both Andrew and Craig when compared to their age-matched peers' data, with shorter visible tongue lengths.

The following sub-sections will discuss these findings in more detail in relation to the current literature. As assessment was required prior to treatment, this section will first of all discuss the strengths and the issues with perceptual assessment. As instrumental assessment was desirable because UVBF was used to try to change articulatory gestures, a comparison of perceptual and instrumental analysis for assessment purposes including the practicalities and drawbacks of both methods will be discussed before discussing the therapeutic design and comparing the clinical implications for VAMS and UVBF. Finally, limitations to the study and future recommendations will be discussed.

5.2 Perceptual Assessment

Speech is acknowledged as one of the key outcome measures of the management of cleft lip and palate (Sell 2005), as the primary aim of therapy is for clients to develop speech perceptually similar (if not indistinguishable) to their peers (Britton et al. 2014). It has been long recognised that perceptual speech analysis is at the core of the profession and is deemed as standard in assessment of speech (Kuehn and Moller 2000). However there remains no consensus over the protocol for perceptually analysing the speech of individuals with CP.

One of the most common methods of assessing the speech in individuals with CP that is reported in the literature is phonetic transcription (Sell 2005; Howard 2011), from which an analysis can be made by completing error pattern analyses, for example in the DEAP phonology assessment (Dodd et al. 2002), or the Phonetic and Phonological Systems Analysis (*PPSA*; Bates and Watson 2012). Further, phonetic transcription allows calculation such as percent consonant correct scores (Shriberg and Kwiatkowski 1982; Shriberg et al. 1986) which can be used to measure the severity of the speech sound disorder and to measure change across a period of time. Heselwood and Howard (2008) suggest that by identifying errors through the analysis of phonetic transcription, therapy targets can be more easily prioritised. Sell (2005) concluded that using blinded independent analysis should be the *gold standard* approach for reporting research and audit of the speech of individuals with CP, recommending that specialist SLTs carry out the analysis due to the complexity of speech disorders associated with CP, however this is not commonly reported in the literature (Lohmander and Olsson 2004). While narrow phonetic transcription offers more information and detail about speech production, it is thought to have limited reliability (Shriberg and Lof 1991; Kent et al. 1999), particularly for complex SSDs associated with CP, and hence intra- and inter-rater reliability should be considered in the analysis of the speech of individuals with CP.

In the current study, both the DEAP Phonology and target specific wordlists were transcribed using narrow phonetic transcription by the tSLT to obtain PCC and

PTCC scores and to complete an in-depth error pattern analysis of the data at each assessment time-point. Intra-rater reliability was carried out three-years post-therapy. The error pattern analysis showed highly variable productions for Craig, and for both children identified typical and atypical speech patterns. Compensatory cleft-type characteristics were identified in the data for both Andrew and Craig. For Andrew, retraction of /n/ to velar placement (throughout all six assessment sessions) and double articulations were perceived (post-therapy) through phonetic transcriptions. For Craig, at baseline /k/ was transcribed mostly as retracted to glottal placement, as was /t/. /g/ was transcribed as either [d] or [n]. Post-therapy, Craig's production of velars were transcribed as correct (90% PTCC). For Andrew, there was no word position effect, whereas for Craig WI was more errorful than WM and WF position and in post-UVBF and maintenance WF had the least amount of errors. For both children, phonetic transcriptions identified cleft type characteristics reported in the literature (Harding and Grunwell 1998; Sell et al. 1999), however for Craig the developmental process of velar fronting was identified for voiced tokens, which is not typically found in children with CP.

When transcribing speech in those with CP, it is important to consider not only the cleft type characteristics, i.e. the compensatory and obligatory errors due to the structural abnormalities in the vocal tract, but also the child's phonological system. Phonological development can be influenced by the articulatory and perceptual constraints presented by a cleft palate (Harding-Bell and Howard 2011). Previous studies investigating phonological development in speakers with CP have identified errors, or patterns, that are both directly related to the effects of CP and/or to typical development of phonology (Hodson et al. 1983; Lynch et al. 1983). It is suggested that those processes which are related to typical development, for example velar fronting, tend to persist longer in children with a CP than those without CP (Chapman and Hardin 1992; Chapman 1993). As a result, assessment protocols for the speech of individuals with CP should incorporate phonological measures (Morris and Ozanne 2003), hence the current study included the DEAP phonology subtest.

Results from the DEAP Phonology subtest identified both types of errors: those which were directly related to CP and those which are typical phonological processes, that have persisted past the developmental stages, for both speakers. Pre-therapy, Andrew presented with variable velar fronting, backing of alveolar plosives and nasals to velar placement, voicing errors, affrication, palatalization and glottal reinforcement. Craig also presented with many phonological processes including velar fronting, post-alveolar fronting, dentalised /s/, backing of alveolar and velar plosives to glottal placement, consonant deletion (initial, medial and final), cluster reduction, stopping and suspected double articulations. However, through perceptual analysis it was difficult to determine the phonetic realisation of the targets, due to complex errors, for both the DEAP phonology subtest and target specific wordlists, which lead to the belief that there may be covert errors. However, due to the study-design, articulatory analysis to detect such errors was not carried out until after therapy had ceased. Had this been carried out prior to therapy starting, covert errors and covert contrasts would have been detected. Harding-Bell and Howard (2011) suggest that phonological contrasts may not always be detected by the human eye or ear and the need for instrumental assessment has been highlighted (Howard 2004). However, in clinical practice, instrumental techniques are not always available and assessment relies solely on perceptual assessment, consequently leading to misdiagnosis of the errors, in-turn potentially leading to the incorrect therapy approach being used.

For both children, phonetic transcription showed an increase in PTCC in untreated wordlists from baseline to maintenance, three months after therapy ceased. In motor-based approaches, it is suggested that a motor-plan has been learned if participants are able to retain and generalise after therapy has ceased (Maas et al. 2008; Preston et al. 2014). For Andrew, PTCC scores increased from 5% at baseline to only 21% at maintenance, showing modest improvement. This is unlikely to represent a clinically significant improvement in Andrew's production of /n/ suggesting that neither therapy was particularly effective for him. However, the GOS.SP.ASS'98 completed nine-months post-study indicated further generalisation of /n/ in WI position, indicating that Andrew had learned the motor-plan for [n] in WI position. In contrast, Craig's PTCC improved from 22% at baseline to 90% at his maintenance recording, suggesting successful integration of velars into untreated words, implying that therapy for Craig was effective. Cleland et al. (2015c) and Cleland et al. (2017b) also found improvement in PTCC scores when targeting velars in children who were velar fronting. However, the therapy target changed for Craig in therapy block two, focusing on /t/, therefore the increase in PTCC in therapy block two (post-UVBF) cannot be attributable to ultrasound.

Results from the perceptual evaluation (see section 3.4) generally corroborate with the PTCC results discussed above, despite previous literature suggesting that transcriptions from single transcribers are unreliable and multi-listener judgements are preferred. The perceptual evaluation sought to determine via a multi-listener perceptual evaluation whether listeners are able to detect an improvement in production of single words at later points in the therapy time-line. To do this, a novel protocol based on a modified two-alternative forced choice design using PRAAT software (Boersma and Weenink 2013) was designed. Overall both children showed improvements from baseline to follow-up with the majority of later time-points being selected as "closer to the adult target". Generally, listeners selected B (later, i.e. post-therapy) as closer to the target when the PTCC score increased. However, due to the differing methodologies of the perceptual evaluation and the phonetic transcriptions it was not possible to correlate results statistically.

The methodology presented in this thesis is a novel protocol and can be compared to other perceptual evaluation methods, such as Visual Analogue Scales (VAS, Munson et al. 2012; Baylis et al. 2015) and Ordinal Scales, such as those found in the CAPS-A (Sell et al. 2009), which have previously been found to be effective tools for measuring speech outcomes in CP (Castick et al. 2017). While evaluations such as the CAPS-A (Sell et al. 2009) evaluates 10 parameters of speech, using an ordinal scale, the current study focuses only on one speech parameter and has only two response choices, with the written target word provided. By focusing on intelligibility/understandability of a specific word, there is no requirement for listeners to be phonetically trained. In fact, the current method has been used with lay listeners for an honours degree student project (Thompson 2015) and for two other student projects with phonetically trained listeners with ease (Alexander 2015; Young 2015). While VAS (e.g., Munson et al. 2012) allow for near differences between tokens, the design in the current study does not. The goal for therapy is for an improvement in understandability and acceptability, essentially for speech to sound "better" post-therapy. The forced-choice design presented here allows listeners to choose from two tokens, taken from two separate sessions (pre- and post-therapy), which one sounds closer to the target, or "better". As it uses the whole word, rather than a single token, it allows for a more real-life evaluation of the understandability of speech outcomes. While the current study uses only single words, the design could be adapted to include connected speech samples which would not be possible using VAS.

The perceptual evaluation provides information on whether listeners select the later time-point as being closer to the target and supplements the PTCC scores. There was a listener trend toward the post-therapy tokens when PTCC scores increased, and a trend toward pre-therapy tokens when PTCC scores decreased (for Andrew post-UVBF). Interestingly, for both speakers the UVBF comparison showed the least amount of significant results. The BL-M showed the highest number of significant results for both speakers, indicating overall improvement in speech outcomes from baseline to maintenance and suggesting generalisation for both speakers. Whilst useful for measuring therapy outcomes in a wider sense by using non-expert listeners for more real-life significance, it does not provide information on the actual errors made by the children and it mostly corroborated with PTCC scores from intra- and inter-rater reliability measures using three listeners, indicating that intra- and inter-reliability was sufficient for analysis in this case and the perceptual evaluation may not have been necessary. In addition, the perceptual evaluation required multiple listeners and asked listeners to attend two one-hour sessions. Whilst quick and easy to use for research purposes, it would not be feasible to carry out such studies in clinical practice. Audits, such as CAPS-A (Sell et al. 2009), or peer review sessions would be more feasible for inter-rater reliability measures within a clinical setting.

Whilst Kuehn and Moller (2000) suggest that perceptual evaluation has the greatest face validity, with perceptual speech assessment considered to be a key outcome measure in the management of CP (Lohmander and Olsson 2004; Sell 2005), in practice phonetic transcription is often subjective, particularly in complex SSDs such as those found in CP, therefore there is a need for further multi-listener perceptual evaluations. The literature suggests using expert listeners for phonetic transcription and perceptual evaluation, with the use of naïve listeners also considered useful to add validity and real-life significance (Bagnall and David 1988; Sell 2005). The perceptual evaluation presented in this thesis included phonetically trained listeners, although none were *expert* in analysing speech of children with CP. Similarly, while the treating clinical and two transcribers conducting inter-rater reliability had experience in transcribing complex SSDs, they had not had any expert training in CP. Despite subjectivities, phonetic transcription is cheap and easy to do in practice (once phonetically trained) and can be done live, without the need for any equipment. However, Sell (2005) suggests that this should be standard practice in any audit or research project for the speech of children with CP to make recordings (audio or video). Sell et al. (2002) investigated the differences in judgements between video and audio recordings of speech of children with CP. They found no statistically significant differences between audio and video analyses, although there was a trend toward video analysis for more accurate and critical ratings of consonant production, hypernasality and nasal turbulence. Despite the preference for video recordings over audio, the nature of the speech-recording medium (i.e. analogue or digital recordings) could impact analysis (John et al. 2003). Gooch et al. (2001) also discuss the importance of the quality of recordings and listening environments and the need to ensure uniformity of amplitude of speech samples. The perceptual evaluation in this thesis did not control for uniformity of the quality of recordings across sessions, which could have accounted for listener bias toward a session of better recording quality.

It is also reported that it is only through recordings that an independent assessment by a specialist SLT, based on randomised order of presentation of recordings, can be carried out (Wyatt et al. 1996). Due to time limitations, an independent specialist SLT assessment was not carried out in the current study. Through interactive discussion between transcribers, this can provide more consistent (and potentially more accurate) results than independent live transcription (Amorosa et al. 1985), however this is time-consuming and is not cost effective for the NHS. Likewise, carrying out a large-scale perceptual evaluation would not be cost effective or feasible within an NHS setting. Storing audio or video data to play back to multiple listeners would also be difficult, with strict Data Protection policies in place in most NHS Health Boards.

This sub-section has discussed some of the strengths and weaknesses identified in perceptual assessment of the speech of Andrew and Craig. The following sub-section will discuss the results from the ultrasound data, which were used to circumvent some of the difficulties with the perceptual analyses above and because UVBF was used for treatment. The findings from the current study will be reviewed in relation to previous literature on CP and EPG and more recently the UTI literature (Bressmann et al. 2011), investigating the compensatory articulations for velar stops in CP speakers.

5.3 Instrumental Assessment

Despite recommendation by the RCSLT (2005) for the use of EPG for speakers with CP, instrumentation plays a relatively minor role in the assessment and treatment of articulation in CP, unlike instrumental assessment of VP function which is more widespread (Gibbon and Lee 2011). As most articulations are hidden from view, instrumental techniques may enhance our knowledge of articulations in addition to perceptual analysis. By using instrumental assessment, objective measurements can be made. Instrumental assessment was also carried in the current study out because ultrasound was used for intervention.

The current study sought to investigate whether articulatory analysis using ultrasound tongue imaging confirmed, disconfirmed or supplemented the phonetic and phonological analyses derived from phonetic transcription and whether improvements in PTCC were confirmed by quantitative measurements, using paired t-tests to identify any differences between two tongue curves (an average of multiple tokens) and how big the difference is by measuring tongue width. Tongue length and image quality was compared to that of age-matched typically developing peers.

Results from the articulatory analysis (section 4.2) showed that phonetic transcriptions were supplemented by ultrasound data, with both covert errors and covert contrasts identified. For Andrew, a covert contrast in /n/ and /ŋ/ minimal pairs was identified. Hewlett (1988) defines the term covert contrast as describing instrumentally measurable differences between perceptually neutralised target phonemes. Most studies investigating covert contrast use acoustic analysis (Kornfeld 1971; Macken and Barton 1980). The phenomenon of covert contrast was originally described by Kornfeld (1971) who found covert contrasts in child speech, specifically in cluster production, through acoustic analysis. Kornfeld found that when two clusters e.g. in “glass” and “grass” were phonetically transcribed as [gwas], spectrographic analysis identified a difference in F2 locus between liquids /ɹ/ and /l/ within the cluster production. Macken and Barton (1980) also identified covert contrast in longitudinal acoustic data of children’s speech. Using acoustic analysis to study the acquisition of stop voicing contrast in WI position in four typically developing children, Macken and Barton proposed three stages in the development of adult-like voice onset time (VOT). Stage one was neutralisation. In stage two they identified covert contrasts and in stage three overt contrasts in VOT. Macken and Barton emphasise that phonetic transcription alone cannot identify these covert errors and highlight the need for instrumental analysis. However, recent studies using VAS suggest that it may be possible to detect covert errors (Schellinger et al. 2017). Whilst most studies investigating covert contrast use acoustic analysis, this does not provide direct information on lingual movements. Visual biofeedback techniques, such as EPG (Hardcastle and Morgan 1982; Friel 1998; Gibbon 1990; Gibbon et al. 1995; Gibbon and Crampin 2001) and ultrasound (Richstsmeier, 2010; McAllister Byun et al. 2016) have also identified covert contrasts, including in speakers with CP (Gibbon and Lee 2017). Whilst phonetic transcriptions of a 36-year-old male speaker, with repaired CP, showed that alveolar and velar targets were transcribed as mid-dorsum palatal stops, EPG data revealed covert contrasts. Namely, the alveolar targets were produced more anteriorly than velar targets. They also found differences in patterns of complete closure in the onset, with complete closure for /t/ being in row four and row eight for /k/. However, the release frames were similar, with contact in the palatal region, which is most likely to have contributed to listeners perceiving alveolar and velars as the same mid-dorsum stop. Using EPG, Gibbon and Crampin (2001) found a reduced separation between alveolar and velar targets; however, a statistically significant contrast was identified in the articulatory data of speakers with CP. Gibbon and Crampin highlight the importance of distinguishing between reduced separation and contrast neutralisation for diagnostic purposes as this may differentiate errors which are phonetic in nature to those that are phonological in nature. This is an important distinction particularly for the therapy used in the current study, which adopted a motor-based approach. If errors are phonological in nature, then a motor-based approach would not be suitable. The quantitative ultrasound measurements in the current study measured whether there is in fact any significant difference (contrasts) and how big the difference is (width), therefore distinguishing between reduced separation and contrast neutralisation.

For Andrew, /n/ was transcribed as [ŋ] (whilst velar nasal targets were transcribed correctly), whilst some double articulations were suspected in later sessions. Perceptually, there was a merger between alveolar and velar nasal targets, therefore errors were categorised into Cleland et al. (2015c) category 1 (identical realisations). However, the ultrasound analysis identified statistically significant differences between the two targets, which were not perceived through phonetic transcriptions, indicative of category 2 (abnormal or underspecified articulations) in Cleland et al. (2015c). Whilst there was dorsal raising for both the alveolar (incorrect) and velar targets in minimal pair data, the analysis showed significant differences in the tongue root, with the suggestion that /n/ was even further retracted than velar placement. This is suggestive of category 3 (non-native sounds) in Cleland et al. (2015c). This is a clear advantage of ultrasound over EPG, which may have also identified a merger between the two consonants due to the palate ending at velar placement. However, without simultaneous ultrasound and EPG recordings it is unknown whether EPG would have detected any covert errors in the velar area. While Cleland et al (2015c)'s categories are based on children with primary SSDs, through Andrew's articulatory analysis, these categories can also be applied to those with CP to provide crucial diagnostic information, which was not identifiable through phonetic transcription alone. The analysis of phonetic transcriptions (see sections 2.2.3 and 2.3.3) suggest a neutralisation between the two nasal targets (category 1 in Cleland et al. 2015c), evidence of a covert contrast indicates that the targets are not completely merged, or neutralised (category 2, Cleland et al. 2015), however the subtle differences are difficult for listeners to perceive due to categorical bias into phonological categories (Liberman et al. 1957). In other words, while homophony may suggest a difficulty at the phonological level, evidence of a covert contrast implies separate phonological representations of two phonemes and not a collapse in contrast. The difficulty therefore lies at a phonetic or articulatory level rather than a phonological level impairment (Gibbon 1999). This has obvious implications for clinical practice. If a collapse in contrast, and therefore a phonological difficulty, is perceived through phonetic transcriptions, the SLT is likely to adopt a phonological therapy approach such as minimal pairs. However, with covert contrasts and consequently a phonetic difficulty present, an articulation or motor-based therapy approach would be more suitable. Although the covert contrasts were not detected until after therapy ceased, a motor-based approach was still adopted for the current study for Andrew, due to cleft-type characteristics and not a typical phonological error. However, an eclectic approach was used during table top activities and phonological activities such as minimal pairs were included.

Another interesting finding for Andrew was the comparison of his incorrect production of /n/ to /t/ and /n/ to /k/ for the DEAP data. Whilst Gibbon et al. (2007) suggest that there is no difference in EPG patterns for alveolar oral and nasal stops in typical adults, they state that results should be used with caution when comparing to paediatric data, particularly those with CP, due to anatomical differences. However, this only investigates alveolar stops and does not look at velar stops. Whilst incorrect, with significant differences found between /n/ and /t/, statistical analysis showed no difference in tongue shape between Andrew's production of /n/ and /k/, which may suggest velar placement for /n/ within the DEAP and therefore suggest that velar oral and nasal stops act in the same way as alveolars. However, there was no comparison of /ŋ/ to /k/ to test whether there were any differences. When comparing the minimal pair data to the DEAP data, qualitatively, with /k/ and /n/ showing identical tongue shapes, but /n/ and /ŋ/ showing a covert contrast, this may also suggest incorrect placement for /ŋ/, with a more anterior tongue root to /n/ and /k/.

In addition to covert contrast, covert errors were also identified in Andrew and Craig's ultrasound data. Covert error occurs when articulatory analysis contradicts phonetic transcriptions and, therefore, identifies the precise nature of a disorder

(Cleland et al. 2017b). Whilst the errors do not give evidence for a contrast, they do provide articulatory evidence of motor based difficulties and similar to the identification of covert contrast will lead to more accurate diagnosis and in turn more appropriate therapy. Various covert errors have been identified in EPG and ultrasound literature. Gibbon (1999) describes undifferentiated lingual gestures in EPG patterns, where difficulty differentiating between coronal and dorsal tongue gestures suggest motor impairment in those children who may present with what initially appear as phonologically disordered patterns. Other errors identified through the use of EPG include misdirected articulatory gestures and, more specifically to cleft palate, double articulations (Hardcastle and Gibbon 2005). More recently, Cleland et al. (2017b) investigated covert contrast and covert error in seven children with persistent velar fronting using UTI. Quantitative ultrasound analysis showed no evidence of covert contrast, however one child showed variable productions for both /t/ and /k/, with retroflexion on both, indicating a SSD at a phonetic level. However, this is not data from individuals with CP, which the evidence from the perceptual evaluation in the current study, along with the existing evidence from the literature, suggests have complex speech errors which are often difficult to detect through transcription alone, resulting in poorer inter-rater reliability.

Gibbon and Wolters (2005) used ultrasound to investigate the speech in an adult with repaired CP, concluding that ultrasound has the potential for investigating abnormal tongue shapes in speakers with CP. However, this was a single-case of an adult and did not investigate the therapeutic application. Similarly, Bressmann et al. (2011) used UTI to investigate the compensatory articulations for velar stops in speakers with CP and found covert articulatory movements, however therapeutic applications were not tested. Compensatory articulations identified in Bressmann et al.'s study show similarities and differences to the current study. Firstly, they identified glottal stops, which were highly prevalent in the analysis of Craig's perceptual data at baseline and pre-VAM. Although these would not be visible on ultrasound data, a lack of lingual movement would suggest a true glottal stop, rather than glottal reinforcement. Craig's data, although phonetically transcribed as glottal stops, indicated covert lingual movements and therefore suggest glottal reinforcement, not glottal stops per se. Bressmann et al. describe velar productions with glottal reinforcement as “glottal and velar co-productions”, which were also identified using UTI for Craig. They also found midpalatal stops, which are also commonly identified in the EPG literature as middorsal palatal stops (Gibbon 2004). These were identified in Andrew’s data, in the example of “pig” and “egg” in sub-section 4.2.1.4. Pharyngeal stops were identified in Bressmann et al. but not in the current study. However, with tongue root retraction for /n/ evident in Andrew’s data, it is uncertain whether this is in the uvular or pharyngeal region.

Through qualitative and quantitative analysis of the ultrasound data, it became apparent that Craig was consistently velar fronting all velars at baseline, which was perceptually not identified. Phonetic transcriptions found that Craig was fronting /ɡ/ to [n] or [d], however /k/ was perceived as [ʔ], with correct productions of [d], but incorrect productions of /t/, which were also transcribed mostly as glottal stops. The DEAP data however showed high levels of variability in the error patterns for both /t/ and /k/, mostly either transcribed as [ʔ] or as unreleased /t/ /k/ with glottal reinforcement. Ultrasound analysis identified lingual movement in the alveolar region for all alveolar and velar plosives. It is well known that retraction to glottal placement is a common compensatory articulation in speakers with CP, as a strategy to maintain a phonological difference (category 3 in Cleland et al. 2015c). However, statistical analysis showed no difference between alveolar and velar tongue shapes in baseline or VAM, indicating a covert lingual merger and thus confirming the phonological process of velar fronting (category 1 in Cleland et al. 2015c). However, as alveolar targets were transcribed as glottal stops, velar fronting was masked by the compensatory articulations, such as glottal reinforcement, or a lingual/glottal double articulation as would be described by Bressmann et al. (2011).

Gibbon also discusses *fronted placement,* which typically involves contact in the palatal region for velar consonants, as an atypical error pattern identified in CP speakers in her 2004 review paper. However, this error is less common in speakers with CP than backing may be. Spriestersbach et al. (1956) suggests that velars are less likely to involve placement errors than alveolar targets, with alveolars typically retracted to velar placement as a compensatory strategy. As previously mentioned, Chapman and Hardin (1992) and Chapman (1993) found that in children with CP typical phonological processes, such as velar fronting, tend to persist for longer than in children without CP, however velar fronting is not a typical process found in speakers with CP. Children should typically have acquired velars by three and a half years old (Dodd 2005). As Craig had surpassed the developmental phase in which velar fronting should have been eliminated, this would suggest a phonological delay.

Previous studies of speech characteristics in speakers with CP using EPG (Morley 1970; Lawrence and Philp, 1975; Golding-Kushner 1995) have also identified overuse of the tongue dorsum. A possible reason for the overuse of tongue dorsum is to aid VP function (Trost 1981). Both Andrew and Craig presented with signs of VPD, such as VP friction, and qualitative ultrasound analysis for both children identified variable overuse of the tongue dorsum in the DEAP and in target specific wordlists. In EPG studies, increased contact has also been identified as an atypical pattern (Gibbon 2004). Although it remains uncertain why there is increased contact in speakers with CP, Hardcastle et al. (1989) proposed that speakers with CP have an impaired development of typical tongue function due to palatal scarring and therefore, have a lack of tactile awareness. Other possible causes could be due to the presence of fistulae, hearing impairment, concomitant verbal dyspraxia or compensatory actions of the tongue apex (Hardcastle et al. 1989). Though ultrasound in the current study was not able to show increased contact as such, because palate traces were not possible, it was able to show overuse of the tongue dorsum. Cleland et al. (2017b) also found undifferentiated lingual gestures, characterised by a whole tongue body movement in a child with CAS. If palate traces were possible and ultrasound images for Andrew and Craig were of better quality, then it may have been possible to measure increased contacts using ultrasound.

Sibilant distortions were also found in Andrew's data. Productions of /s/ and /ʃ/ were transcribed as a palatal fricative; however, this was variable. Palatalisation of fricatives is previously reported in the EPG literature (Michi et al. 1990; Yamashita and Michi 1991; Howard, 1998; Howard and Pickstone 1995). It is suggested that increased contact on EPG patterns, or overuse of the tongue dorsum, can result in reduced control of the lateral margins of the tongue (Gibbon 2004), which is essential for speech production (Stone et al. 1992). As ultrasound data in the current study was recorded in the midsagittal view, information of the lateral margins of the tongue is missing. It is possible to use ultrasound in the coronal view, which would show lateral bracing. For example, Bressmann et al. (2010) propose quantitative measures for coronal ultrasound (total distance travelled and concavity). Another way to overcome this would be to record simultaneous ultrasound and EPG.

5.3.1 Practicalities of Using Ultrasound

By comparing the errors identified using ultrasound in the current study to those identified in previous EPG and ultrasound studies, this allows for a comparison of EPG and ultrasound as assessment tools for individuals with CP. Whilst EPG is recommended by the RCSLT and is able to identify many of the compensatory articulations noted in the literature, it has its drawbacks. Where EPG can only identify errors, or patterns, as far back as the velum, ultrasound is able to identify errors in the uvular and pharyngeal region. With EPG, this would present as an open pattern, not providing any lingual information. Ultrasound is also useful for identifying lingual errors in vowels, whereas EPG would provide limited information. Ultrasound also has the advantage that there is no requirement for individualised hardware, whereas each EPG palate costs around £500 to make making the use of EPG expensive for assessment purposes. However, ultrasound also has its limitations for assessment.

Whilst EPG provides details on the amount of tongue-palate contact, ultrasound only provides data on the tongue surface. In midsagittal view, ultrasound data provides no information on the lateral margins of the tongue and at present there are no quantitative measurements for ultrasound data collected in the coronal view using AAA, although Bressmann et al. (2010) propose quantitative measurements for coronal ultrasound, such as *total distance travelled* and *concavity*. For some speakers, the length of the surface of the tongue that is imageable using ultrasound is short due to large hyoid and mandible shadows. This speaks true of both Andrew and Craig, who when compared to age-matched peers had shorter tongue lengths than expected. It is likely that this is because both speakers had a short space between the hyoid bone and mandible due to small mandibles (small chin). Studies have found structural abnormalities to be a feature in those with cleft palate, such as asymmetry in the face (Bugaighis et al. 2010) and asymmetry in the palate and oral cavity

(Kilpelainen and Laine-Avala 1996). However, asymmetry of the mandible does not appear to be an associated feature (Kurt et al. 2010), which would be the obvious feature that may impact on an ultrasound image. Andrew presented with Hemifacial Microsomia, a syndrome associated with cleft palate and facial asymmetry. As a result, the headset used for probe stabilisation was difficult to fit symmetrically. In turn, the ultrasound probe was not always in the midsagittal position, presenting artefacts or skewed images on the ultrasound data (see chapter 4). Although Craig did not present with facial asymmetry, his head was small and there was a lot of headset movement resulting in variability in the ultrasound images. The headset used in this study is also primarily made for adult speakers; therefore, a smaller headset would have been beneficial and may have prevented as much probe movement. Zharkova (2013) proposes quantitative measurements that do not require headset stabilisation. However, these measurements do require equal hyoid and mandible shadows. Due to the poor quality of the images for the speakers presented in this thesis, with a smaller portion of visible tongue than TD peers (which Zharkova's measurements are based on), it was not possible to carry out these measurements on the data for this thesis. Due to these potential disadvantages of using ultrasound for individuals with CP, it is beneficial to consider the anatomical features that may be associated with CP and their suitability when considering ultrasound as an assessment tool.

From a diagnostic point of view, the ultrasound analysis provided crucial information, such as covert error and covert contrast. Whilst it was clear from qualitative results where there were similarities and differences in tongue shapes, statistical analysis and quantitative measures were key, for identifying subtle differences between tongue shapes and for measuring the visible tongue length and tongue width. These analyses were carried out after therapy had ceased due to time limits between blocks of therapy. Although the raw images were interpreted live during therapy sessions and qualitative measures were carried out during the period of time participants were enrolled on the study, the crucial information identified through quantitative measures that would have been useful diagnostically was missed during the treatment process. These subtle differences are not easily interpreted from live, real-time or recorded raw ultrasound images and quantitative analysis is a necessity for identifying such subtle errors. Clinically, this would require additional preparation, recording and analysis time making ultrasound a timely and potentially costly tool. This highlights the need for automatic tracking of ultrasound data, for clinical applications, to allow ultrasound to become a readily available tool for clinicians to use quickly and effectively for assessment purposes. The ULTRAX2020 Project (2017-2020), a follow-up project to ULTRAX (2011-2014), aims to do just that, by developing a method of classifying tongue shapes to form the basis of an automatic assessment and objective therapy outcome measures.

5.3.2 Summary of Instrumental Analysis

This subsection has discussed the benefits of including instrumental analysis in the assessment of speech characteristics associated with CP, providing information on covert errors and covert contrast that were not identified through perceptual assessment. While compensatory error patterns found in speakers with CP are reportedly adopted to facilitate phonological development, the instrumental assessment here suggests that compensatory articulations in both speakers are in fact a result of incorrect motor plans. Covert lingual mergers found in Craig's data were indicative of category 1 suggested by Cleland et al. (2015c), with no statistical differences in the tongue shapes for /t/ and /k/. For Andrew, covert contrasts were indicative of category 2 and errors such as palatalisation, uvularisation and glottalisation in both speakers are indicative of category 3 (Cleland et al. 2015c).

These findings are crucial, suggesting that instrumental assessment is essential diagnostically for speakers with CP due to the complexity of their speech. However, these errors would not have been identified through live interpretation or qualitative measurements alone, thus highlighting the need to record data and obtain quantitative measures. This also allows data to be compared to typically developing children, which for Andrew and Craig identified a poorer image quality and shorter length of visible tongue than age-matched peers.

Now that the assessment findings have been discussed in relation to the existing literature and clinical implications, the following sub-sections will address the therapeutic application of ultrasound visual biofeedback and visual articulatory models, including the therapeutic design which will be addressed first.

5.4 Therapeutic Design

The therapy study used a single-subject multiple baseline design with alternating treatments. Andrew and Craig received six assessment sessions and two blocks of motor-based therapy, each containing eight one-hour therapy sessions using either Speech Trainer 3D or ultrasound. The first block of therapy used Speech trainer 3D and it was hypothesised that PTCC scores would remain stable due to little or no effect of VAMS previously noted in the literature (Fagel and Madany 2008; Massaro et al. 2008; Cleland et al in press). However, this was not the case and both children showed improvement in PTCC scores post-therapy block one (VAMs). As reported in chapter 2, various other methodologies were considered for the therapy design. However, with limited numbers and constraints on the allocated time to complete a PhD, larger scale studies with a cross over design were not possible. A design that may have been suitable is the randomised block design, such as that in Sjolie (2015). Similar to an alternating treatment design, the randomised block design presents treatment order randomly instead of it remaining the same. The current study presented two treatments (VAMs and UVBF) in an ABACA design. If it were to use the randomised block design, Andrew and Craig would have randomly been allocated VAMs or UTI for their first block of therapy. In the current design, speakers underwent two assessment sessions prior to starting therapy (baseline and pre-VAM), although these sessions were only one week apart leaving limited time between baselines. Ideally, more than two baseline assessments would have been included across a longer period of time, to ensure stability prior to beginning therapy. A motor-based therapy approach was adopted for both children, similar to that in Preston et al. (2014) and Cleland et al. (2015c). This method has been used extensively to treat those with speech sound errors persisting past the typical age of acquisition (i.e. beyond 8-9 years; Shriberg et al. 1997; Preston and Edwards 2007). Due to the resultant structural abnormalities in speakers with CP, children are at a high risk of developing speech difficulties (Vallino-Napoli 2011). Whilst this is routinely managed surgically, Hardin-Jones and Jones (2005) report that the majority of pre-schoolers with palatal repairs (68% of 212 preschool-aged children) still require therapy focused on improving their speech and speech difficulties may persist beyond childhood into adulthood, with persistent compensatory articulations resulting in incorrect motor plans. Despite this, there is a lack of agreement on the best type of intervention for treating articulatory errors associated with CP, with a recent systematic review finding little evidence to support any particular technique (Bessell et al. 2013). However, of the 17 studies that did meet inclusion criteria, 10 report the results of motor based approaches, suggesting that the professional opinion is that interventions which employ on the principles of motor-learning may be appropriate for individuals with CP (Ruscello and Vallino 2014).

Chapter 1 discusses the principles of motor-learning, which can be applied to motor-based therapy and were applied to the therapy protocol in the current study (see Table 9). Firstly, it is important to consider the two phases of motor-based intervention: pre-practice and practice. Pre-practice is intended to prepare the speaker for practice sessions (Schmidt and Lee 2005). Maas et al. (2008) suggest that the goals of pre-practice are to ensure that speakers have a motivation to learn, have an adequate understanding of the therapy task (including what responses are considered correct) and to ensure that their target sound is stimulable. They also note the importance of ensuring adequate auditory-perception abilities. Pre-practice is essentially where participants will acquire a new motor-skill. Pre-practice and practice were included in both blocks of therapy within the current study. The goals of pre-practice were addressed by using Speech trainer 3D or UVBF to model the correct placement for the target with an anatomical reference (e.g. the soft palate for velars or the alveolar ridge for /n/) in order for acquisition. During acquisition, three methods of practice fraction were considered: simplification, segmentation and shaping. Through shaping, facilitative vowels or consonants were used to increased learning, e.g. /o/ was used to teach /k/, along with gestural instruction from the treating SLT.

By using an iPad, for which the children were already familiar and keen to use, it was believed that this increased their motivation to learn. While it is suggested that VAMs are not the key agent of change in learning new articulations (Cleland and Scobbie in press), both children made more progress during therapy block one using Speech Trainer 3D. Despite this, both children selected ultrasound as their preferable tool because they were able to see their own tongue moving. It was felt that for both children, motivation levels in therapy block two had decreased and that, particularly for Andrew, they did not have an adequate understanding of how to interpret the ultrasound image. This was not helped by the poor image quality of both of the children's ultrasound data and the lack of tongue tip image due to a large mandible shadow when aiming to target alveolar sounds. Retention occurred during the practice phase of intervention, through using both Speech Trainer 3D and UVBF and generalisation occurred through table-top activities without the use of visual feedback tools. Evidence of generalisation was found in the GOS.SP.ASS assessments for both children post-therapy.

Secondly, it is important to consider the practice conditions on motor learning (Maas et al. 2008). It is suggested that using a large number of practice trials or sessions will increase motor learning, which should be practiced over a prolonged period of time. In terms of practice variability, Preston et al. (2014) suggest that constant practice is useful when initially acquiring a new motor skill (i.e. speech sound). However, in order to achieve retention and generalisation, i.e. true learning, therapy should shift toward using variable practice, including linguistic variability i.e. practicing the therapy target in different contexts within a syllable, rather than in the same syllable context. In the current study, practice was distributed over a period of nine months (16 once weekly therapy sessions) with a large number of trials per session, however this was not measured directly and a set-number of trials was not included within the therapy protocol. Practice variability was both constant and variable, depending on the complexity of the tasks. Targets were practiced in various syllable or word positions in a variety of different vowel environments, with increasing practice complexity, similar to that in Cleland et al. (2017c). Another suggestion regarding the practice schedule is to use a randomised schedule, i.e. mixing therapy targets for example targeting /k/ and /t/ within the same session. The current study used both a blocked practice schedule (i.e. different targets were practiced separately) and a randomised schedule (i.e. therapy block two for Craig which targeted both /t/ and /k/ to maximise contrast). Maas et al. (2008) also discuss attentional focus, which can be internal (a focus on bodily movements such as articulatory placement) or external (a focus on the effects of the movements such as the acoustic signal). In the current study, both internal and external attentional focus were used. By using ultrasound biofeedback, Andrew and Craig were able to view their own tongue in real-time to focus on their own articulatory placement. This can also be applied to the Speech Trainer 3D app by focusing on where the tongue is in relation to the passive articulators, which cannot be viewed on ultrasound. This increased acquisition and retention of the targets, while table top activities increased generalisation. It is suggested that errorless practice supports accurate acquisition of a new motor skills however for children with CP this may not always be possible due to any anatomical abnormalities. The current study used errorful practice, which provided opportunity for mistakes, in turn providing a chance for both Andrew and Craig to identify their own errors and self-correct (Maas et al. 2008; Bergan 2010). By using Speech Trainer 3D and UVBF, Andrew and Craig were able to identify any errors made by either scrolling through the app on the iPad to identify the tongue shape they felt they were making and then playing the video for the correct articulation in order to self-correct, or by using the biofeedback elements of ultrasound (live biofeedback or delayed biofeedback, i.e. watching recorded videos of their own articulations).

As well as practice conditions, it is also important to consider the different types of feedback given throughout a session, how often feedback should be provided and the timing of feedback. Preston et al. (2014) suggest that in order to establish a new motor skill, participants must acquire and learn their skill and that acquisition can be enhanced by high-frequency feedback on how the speech sound is produced (knowledge of performance, KP feedback) and judgement of correctness (knowledge of results, KR feedback). In the case of UVBF, KP is provided by the ultrasound image and is used by both the speaker undergoing therapy and by the tSLT. The tSLT may also provide verbal KP feedback. Although KP using biofeedback was not possible during therapy block one using Speech Trainer 3D, KR and verbal KP feedback was possible, for example by providing information on the learned gestures, using Speech Trainer 3D to provide a visual reference for instructions. Preston et al. (2014) also note that feedback should be randomised as to which attempts (or trials) should receive feedback, to reduce the frequency of feedback, and that feedback should be changed in nature from KP (with or without KR) to only KR, in order for the speaker to retain and to generalise a new motor skill. In the current study, half of each session used the visual feedback tools and the other half of the session used a more traditional approach, incorporating minimal pairs and other table top tasks, in order for Craig and Andrew to retain and generalise their therapy target. Feedback frequency was high in earlier sessions of therapy, which reduced to low frequency in later sessions in order to increase motor learning. The timing of feedback was concurrent, immediate and delayed, by playing recorded UVBF videos back to Andrew and Craig. Both concurrent and immediate feedback improves performance during practice; however, delayed feedback is believed to be more helpful for learning as the delay encourages learners to detect any errors and self-correct, therefore improving on their next attempt (Ballard et al. 2010; Murray et al. 2015; McLeod and Baker 2017). Using the bespoke therapy version of Articulate Assistant Advance (AAA; Articulate Instruments 2012) ultrasound software, recordings were made of Andrew and Craig's attempts of their therapy targets (/n/, velars or /t/. By using delayed feedback (not to be confused with delayed auditory feedback), this allowed Andrew and Craig to use their internal feedback system to watch and rate their own productions, in turn self-correcting and improving their next attempt. It also provided opportunities for discussion between the SLT, participant and parent.

Andrew acquired the motor plan for /n/, but did not retain or generalise this into untreated words suggesting that he did not truly learn the motor plan for /n/. However, the post-therapy GOS.SP.ASS'98 (Sell et al. 1999), completed nine-months post-study, suggests generalisation of /n/ into syllable initial, word initial placement. Craig, on the other hand acquired, retained and generalised into untreated words, suggesting that he did learn the motor skills for velar plosives, but he was not able to generalise for /t/, the treatment target in therapy block two using ultrasound. However, GOS.SP.ASS'98 results, completed by the CLP specialist SLT nine months post-study would suggest that Craig had generalised and maintained the lingual placement for all alveolar and velar plosives, thus suggesting that therapy reinforced the contrast between alveolars and velars.

5.5 Visual Articulatory Models and Ultrasound Biofeedback

The primary aim of the current study was to test the effectiveness of ultrasound visual biofeedback and visual articulatory models, to determine whether children were able to acquire, retain and generalise therapy targets by using offline models or whether they additionally require the biofeedback element ultrasound offers. In the most general sense, when using visual biofeedback, individuals learn to self-regulate and change an automatic physiological function of which they would not otherwise be aware of, in a positive direction, through monitoring (France and DeAngelo 2016). In the case of UVBF in the treatment of SSDs, this is not necessarily the case as we have conscious control over our articulators, even though these are largely hidden to the human eye. Despite the tongue being largely hidden in the vocal tract, and in the absence of other visual cues, people will still acquire speech (Cleland et al. 2013). According to the Motor Theory of Speech Perception, listeners will use their own articulatory knowledge of phonetic gestures at a subconscious level to perceive speech (Liberman and Mattingly 1985; Cleland et al. 2013). By using VBF tools, such as UTI, speakers are able to see their lingual movements in real-time and "read" their images, increasing their knowledge of phonetic gestures and applying the new knowledge to both perception and production of a new articulatory target. It is suggested that being able to see the lips move during speech enhances perception (Benoit and Le Goff 1998), with models of speech perception/production, such as the McGurk effect, showing that speakers have the ability to implicitly learn how to integrate lip information into the perceptual system (McGurk and MacDonald 1976). Similar to speech reading, Badin et al. (2010) suggest that speakers are able to implicitly tongue read a VAM. Indeed, both Craig and Andrew were able to acquire new articulatory gestures using Speech Trainer 3D, indicating their ability to read the VAM in the iPad app. Previous studies (Treielle et al. 2014; Cross et al. 2006) provide evidence that using UVBF activates the areas of the brain responsible for generating mirroring activity. Crucially, for therapy using UVBF, when shown the correct production of a new articulatory movement, the child's mirror neurons will likely trigger the imitation system to perform the movement themselves, in turn being able to self-regulate and change their own articulations in real-time (Cleland

and Scobbie in press). Although videos of articulatory gestures were demonstrated through Speech Trainer 3D, it is unknown whether these are based on anatomically correct data, therefore it cannot be assumed that these are correct models. By using a bespoke version of AAA (Articulate Instruments 2012), it was possible to show Andrew and Craig videos of correct articulations from age-matched typically developing peers, in order for them to manipulate their tongue in real-time to imitate the tongue movements in the videos. During therapy using UVBF, it was also possible to record and play back the speaker's productions to them, allowing discussion between the SLT and participants and allowing the SLT the provide accurate feedback to participants to improve their next attempt. When referring back to the conditions of feedback in motor-learning, delayed feedback is thought to be more useful for learning a new gesture during the practice (Ballard et al. 2010; Murray et al. 2015; McLeod and Baker 2017). This was a key part of the therapy protocol in the current study, which has not been reported in previous studies using UVBF.

UVBF has become increasingly popular over the past few years, with around 30 small number studies (see Table 5). The majority of the studies have reported on developmental SSDs (Shawker and Sonies 1985; Adler-Bock et al 2007; Bernhardt et al. 2008; Modha et al. 2008; Klein et al. 2013; Lipetz and Bernhardt 2013; McAllister Byun et al. 2014; Preston et al. 2014; Cleland et al. 2015c; Hitchcock and McAllister Byun 2015; Lee et al. 2015; Bressmann et al. 2016; Heng et al. 2016; Melo et al. 2016; Preston et al. 2017a; Sjolie et al. 2016) other studies have included a range of client groups, including CAS (Preston et al. 2013; Preston et al. 2016a; Preston et al. 2016b; Preston et al. 2017b), Down's Syndrome (Fawcett et al. 2008), Hearing Impairment (Bernhardt et al. 2003; Berndhardt et al. 2005a; Bacsfalvi et al. 2007; Bacsfalvi 2010; Bacsfalvi and Bernhardt 2011), Glossectomy (Blyth et al. 2016) and Cleft Palate (Roxburgh et al. 2016). Nineteen of these studies have reported improvement in speech outcomes, whilst 11 have reported mixed results. Most of these studies have used only UVBF, with the exception of Bacsfalvi's studies comparing ultrasound and EPG, and Sjolie et al. (2016) which compared two types of intervention (ultrasound and no ultrasound) in a blocked randomised design to treat /r/. Sjolie et al. (2016) found that two out of four children showed no acquisition, retention or generalisation of rhotics, with only one participant showing a significant advantage of ultrasound over no ultrasound for acquisition. There were no differences between treatment conditions for retention or generalisation. Similarly, Preston et al. (2016b) evaluated the acquisition and generalisation of rhotics in three children with CAS, over 14 sessions with UVBF. While participants showed increased accuracy during treatment, none of the children demonstrated generalisation into untreated words post-therapy. Cleland et al. (2015b) also found no advantage for ultrasound verses articulatory teaching of novel speech sound for 30 typically developing children. In Cleland's study both types of teaching were equally successful, however, the authors conclude that it is perhaps the *persistent* nature of SSDs that makes VBF a necessary approach. Indeed, most children in clinical studies of UVBF (and EPG) present with a history of failure to acquire particular articulations despite sometimes years of therapy (Wood and Scobbie 2003; Carter and Edwards 2004; Preston et al. 2014; Cleland et al. 2015c). This may also be similar for those children with CP who present with compensatory articulations. While the literature suggests that the compensatory error patterns found in speakers with CP are adopted to facilitate phonological development, it could be argued that compensatory articulations could also a result of incorrect motor plans if they persist post-surgery. Preston et al. (2014) suggest that inappropriate phonetic realisations, or compensatory articulations in the case of CP, occur due to an inappropriate motor plan. Due to this persistent nature of compensatory articulations, despite surgical interventions to improve any VPD, VBF techniques such as UVBF or EPG may be necessary for the acquisition of a new articulation. Indeed, Gibbon and Wood (2010) suggest that VBF is most useful for establishing motor programmes for new articulations (i.e. sounds that are non-stimulable).

The current study used both UVBF to VAMs, which has not been previously reported in the literature (with the exception of Roxburgh et al. 2016, which reports the findings from the perceptual evaluation of this thesis). Similar to Sjolie et al. (2016) and Cleland et al. (2015c), there was no clear advantage of using UVBF, with PTCC scores increasing more with VAMs than post-UVBF. This was contrary to expectations that PTCC scores would remain stable through therapy block one of VAMs, due to the lack of biofeedback. This raises questions about whether the two children in the current study had truly persistent SSDs, or whether they were likely set to benefit from any motor-based approach. Indeed, Bessell et al. (2013) suggest that motor-based approaches are effective in treating the speech of individuals with CP. Andrew's error in his production of /n/ (backed to [ŋ]) was persistent and he had received unsuccessful therapy to target this sound previously. The community SLT had been targeting Andrew's speech sounds, from age 3;0. However, therapy was less frequent between age 5;8 and referral to the project, perhaps suggesting that the dosage of therapy provided was not adequate. In contrast, Craig's production of velars had not been previously treated and thus does not truly constitute a "persistent" SSD. Craig made rapid progress in block one using Speech Trainer 3D (19% pre-therapy; 72% post-therapy) with little further improvement in block two with ultrasound (78% pre-therapy; 88% post-therapy).

Andrew had a mild unilateral hearing loss, which would perhaps imply that he would improve more using ultrasound than using Speech Trainer 3D, since UVBF circumvents some of the need to monitor productions through auditory feedback. However, Andrew had difficulty manipulating his tongue shape to match a static image of his own target alveolar tongue shape to produce /n/. Clinical notes suggest that Andrew was only starting to make progress toward the end of his block of therapy, achieving around 60% [n] correct in *treated* single words containing /n/ in WI position during session seven, and may have benefitted from more sessions to achieve his target tongue shape, thus suggesting that future studies need to take the dosage of UVBF into account. Cleland et al. (2015c) showed improvements in speech post-therapy in 12 sessions of ultrasound therapy for 6/7 children, however, one child made no progress after the first block of therapy, but went on to do so after a further block. The present study provided 16 sessions overall, however only eight of the sessions were using ultrasound, again highlighting the need to investigate dosage.

Clearly it would be premature to conclude that UVBF is not a useful approach for children with CLP as the current study had a limited number of participants. Nevertheless, the success of particularly Craig with the VAM highlights the need to take this type of technology seriously. Mobile apps, particularly those on iDevices

(Apple 2012) are becoming increasingly popular and are now a key tool for Speech and Language Therapists (Gosnell 2011). A 2011 document published by the American Speech-Language Hearing Association lists the advantages and disadvantages of using iPads as a therapy tool. Advantages include better communication through the use of e-mail, cost and time savings, progress monitoring, adaptability, and motivation. Disadvantages include initial investment on the purchase of the device and connectivity. Although the initial cost of the device is listed as a disadvantage by the American Speech and Hearing Association (ASHA 2017), this is significantly less expensive than the initial cost of an ultrasound machine.

In 2011, at least one in five SLT clients used a handheld device (Dunham 2011), with 8% of children under eight years in America owning a tablet device (Common Sense Media 2013). In 2013, 40% of children in America owned a tablet device and 63% of families were using smartphones (Common Sense Media 2013). Both Andrew and Craig were able to navigate their own way around using the app during therapy and download the app to their own devices for home practice, due to it being commercially available at a low cost of £7.99. However, the app did have some usability issues. While it provides a visual articulatory model of American-English consonants and vowels it is not based on anatomical data and is therefore not anatomically correct and for certain speech sounds, e.g. for [k], the video and audio data were not synchronised. Both children independently commented on this. Although Speech Trainer 3D lacks in anatomical accuracy, it provides a child-friendly model for tongue movements in relation to other areas of the vocal tract and acts as a useful control to ultrasound tongue imaging and it is unknown whether anatomical accuracy is crucial for the clinical application of such models. Similarly, EPG is also not anatomically correct, however there is a large evidence base for its effectiveness in treating SSDs in individuals with CP. Further, Speech trainer 3D provides fuller anatomical information than ultrasound, by picturing the entire vocal tract, rather than simply tongue movement.

Although iPads may be cheaper to buy than ultrasound and ASHA (2017) states that an advantage of iPads is client motivation, both Craig and Andrew reported that they preferred using ultrasound because they could see their own tongue moving in real-time. Although cheaper to buy, this does not necessarily mean that Speech trainer 3D would be more cost effective. For VAMs to be more cost effective than UVBF, they would also have to require less dosage and have a faster differential rate of change than ultrasound. Although in both blocks of therapy, they moved up to sentence level, this was not measured with a step-up level of 80% accuracy, which makes it difficult to measure accuracy for differential rate of change. The question also still remains about what actually makes visual biofeedback work and whether speakers require the elements of biofeedback (real-time and delayed biofeedback) or if seeing a VAM is enough. Although there is limited evidence for the use of VAMs (Fagel and Madany 2008; Massaro et al. 2008; Cleland and Scobbie in press), previous studies use only a VAM for speech reading and do not include the use of VAMs as an adjunct to motor-based therapy with the addition of explicit instruction and SLT feedback, such as that presented in the current study. Using the Speech trainer 3D app allowed participants to scroll through and watch videos of articulatory gestures independently, however also allowed the SLT to provide gestural instruction and auditory KP and KR feedback. UVBF may have an advantage over VAMs, with the additional biofeedback elements such as the child seeing their tongue moving in real-time, triggering mirror neurons for self-regulation (Cleland and Scobbie in press). The other advantage of UVBF may be the delayed feedback of being able to watch recordings of attempts at a target. Using UVBF, this also allows the treating SLT to provide more accurate feedback and gestural instruction during therapy. For both Andrew and Craig, a greater rate of change was found in therapy block one using Speech Trainer 3D, therefore, it would be important to consider why this may be the case.

When using ultrasound visual biofeedback, speakers undergoing treatment are faced with the new task of interpreting the images and using these images to monitor and manipulate their own tongue movements in real-time to achieve a new articulation, which faces the issue of their ability to tongue-read. A small number of studies have attempted to investigate whether listeners have an intuitive ability to tongue-read, using various forms of visual articulatory models such as those discussed in sub-section 1.5.3. Mostly used for pronunciation training in second language learning, VAMS (or “talking heads”) provide an artificial articulatory animation, based on

MRI or EMA data (Kröger 2003, Badin and Serrurier 2006; Badin et al. 2010), in a midsagittal view. However, very few studies investigate their effectiveness for this. Cleland et al. (2013) investigated whether there was an intuitive ability to read EPG and ultrasound images and whether participants were able to interpret images from one tool more so than the other. They found that participants scored above chance in both the ultrasound and EPG conditions, confirming that images from both techniques can be interpreted intuitively to some degree. Participants were able to identify consonants in both the conditions and vowels in the EPG condition, which is surprising as there is limited tongue-palate contact with vowels and vowels are more easily viewed on ultrasound.

In the current study, both speakers made improvements in PTCC using the VAM, Speech Trainer 3D, and both speakers were able to label sounds phonetically within one session indicating their ability to tongue-read. However, the reason behind their ability to tongue-read was not directly measured. Craig, interestingly reported, that ultrasound was easier to "read" than the visual articulatory model; although it is unknown what he meant by "read" in this case. Future studies should consider testing directly the speaker's ability to tongue-read.

In the literature on speech reading, or silent lip reading, Kyle et al. (2013) propose a method for testing the ability to speech read in children using the Test of Child Speechreading (ToCS), a computer-based test that measures child speech reading at three psycholinguistic levels: (a) Words, (b) Sentences, and (c) Short Stories. This method for assessing speech reading could be adapted for use with a visual articulatory model to assess tongue-reading at the same three psycholinguistic levels. Speech-reading differs from tongue-reading in the sense that speech reading involves face-to-face communications, whereas tongue-reading of ultrasound images involves interpreting images in a midsagittal view. Commonly though, both speech-reading and tongue-reading are multimodal and speech perception in both cases involves both auditory and visual information.

Visemes are clearly an important component in speech perception although there is no note in the literature of visemes including tongue movements, due to the fact these are largely hidden, minus the tongue tip movement for interdental sounds for example. Typically, the speech perception model (Figure 98) involves a speech signal, including an auditory cue (phoneme) and a visual cue (viseme) which in turn leads the listener to perceive the spoken word.

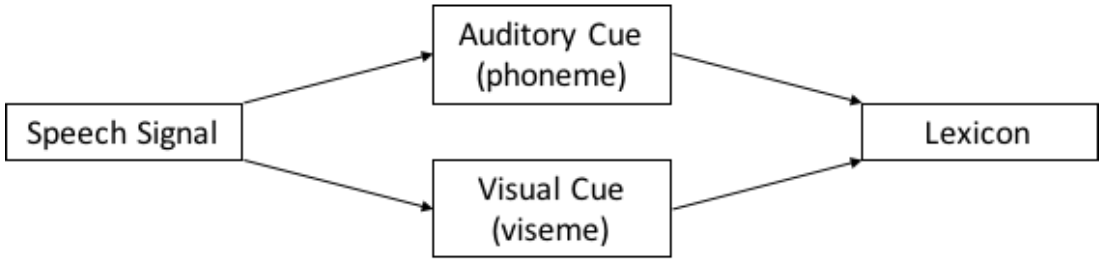


Figure 98 Speech Perception Model (Adapted from Peelle and Sommers 2015)

Typically, a viseme, is described as an articulatory gesture such as the lips moving together, the jaw moving or teeth exposure, derived from a group of phonemes with the same visual appearance (Auer and Bernstein 1997). In the current study, a lingual viseme, that is not easily viewable without a biofeedback technique, was introduced through the use of UVBF. This enables speakers to not only tongue-read an articulatory model, but also to monitor and manipulate their lingual gestures in real-time. EPG has been used to provide visual biofeedback to treat speech sound disorders (Bernhardt et al. 2003) with a large number of single-case and small group studies showing positive results, indicating that speakers are able to read the images of tongue-palate contact. It is said that EPG is relatively intuitive to understand (Gibbon and Wood 2010), even for those who may present with a cognitive impairment (Cleland et al. 2009). However, like UVBF there has been little exploration of why it is useful for the speaker undergoing treatment to view their own articulation and precisely how being able to see their own tongue, or tongue-palate contact, is able to help improve articulatory errors that have not resolved through other therapy approaches.

The current study suggests that participants have the capacity to tongue-read both ultrasound and visual articulatory models, in order to improve lingual errors, with a higher percentage of improvement found in the block of therapy using Speech Trainer 3D. Contrary to expectations, both Andrew and Craig were able to label a speech sound and produce their therapy target within one session, albeit inconsistently. This may be due to the fact that the visual articulatory model presents a context for lingual patterns and is less abstract than a raw ultrasound image, however this was not measured directly. This returns to the question about whether biofeedback is necessary in the remediation of speech sound disorders associated with cleft palate. The current study does not provide sufficient evidence to support the need for biofeedback. While PTCC scores and the perceptual evaluation showed a further improvement in Craig's speech after ultrasound this is to be interpreted with caution because he had clearly acquired the new speech sound before commencing the UVBF therapy (Roxburgh et al. 2016). Indeed, VBF is thought to be most useful for establishing motor programmes for new articulations (Gibbon and Wood 2010), thus probably rendering it unnecessary once Craig had learned to produce a velar articulation in the VAM block of therapy. It is likely that the further improvement he achieved would have been possible without ultrasound. For Andrew, listener judgements and PTCC scores unexpectedly indicated a deterioration in production of his therapy target /n/ post-therapy using UVBF. This highlights the need for future studies investigating how exactly biofeedback works. The current study does not compare visual feedback techniques to traditional therapy approaches, such as a motor-based therapy approach. Therefore, there is a need for a larger scale study comparing ultrasound with visual articulatory models, EPG and traditional therapy approaches. It did however investigate and compare the clinical application of ultrasound and VAMs, which has not yet been investigated in the literature.

5.6 Limitations and Future Implications

Whilst there are many strengths in the current study, this chapter has also highlighted limitations in the design. Firstly, whilst this study intended to recruit six children, only two children took part. Both children presented with a repaired submucous cleft palate, therefore other, potentially more severe, cleft types were not included. Although PTCC scores indicated improvement overall for both children, results from two children, who presented with different error types, cannot account for all individuals with CP. Future studies would benefit from including a larger number of participants with a range of cleft types, to investigate the effectiveness of ultrasound for both assessment and therapy. In fact, there is an ongoing study, currently investigating the use of ultrasound for diagnostic purposes in CP speakers (Cleland et al. 2017a). Both Andrew and Craig presented with poorer image quality than their age-matched peers, however this may have been coincidental. It is clear from the typically developing data that images vary in quality; however, Craig's images in particular were of poorer quality than the "worst" of the typically developing age-matched peers. This does however, raise clinical considerations of cleft-types and any other associated facial abnormalities such as small jaw or facial asymmetry which are commonly associated with CP and whether participants presenting with additional structural anomalies may not be suitable for ultrasound therapy. Cleland (2017-2018) will hopefully be able to answer some of these questions, which were not possible to answer in the current study due to small numbers.

Secondly, due to small numbers, a group design or a randomised block design, was not possible for the current study. Although the single-subject multiple-baseline design with ABACA alternating treatments design was a suitable methodology for this study, it would have helpful to compare UVBF and VAMs more rigorously by having some participants receive UVBF first, rather than VAMs, or by having one group of children receive therapy only with VAMs and one group of children receiving therapy only with UVBF for comparison of group results. Whilst results could not conclude that ultrasound showed a clear advantage over VAMs, results may have been different had they received UVBF first, since UVBF is said to be most useful for acquiring new sounds and therefore may not have been the most suitable tool to use in therapy block two for Craig or Andrew, with PTCC scores showing retention and generalisation (although not significant for Andrew) after therapy block one with VAMs.

This leads on to a third, and important, limitation in the design of speech materials for measuring these therapy outcomes. There are obvious gaps in the wordlists for both Craig and Andrew, with no treated wordlists being recorded pre-therapy to assess retention within treated words. Although, modifying the wordlists throughout therapy allowed the targets to be tailored specific to their needs at the time, rather than prior to therapy starting. For Craig in particular, the therapy target changed during therapy block two, and therefore it was not anticipated that another wordlist for /t/ would need to be recorded. In fact, therapy targets for Craig were an issue in themselves. While the untreated velar wordlist showed increase in PTCC post-UVBF, the target for Craig post-UVBF focused mostly on /t/, with velar targets being worked on during table-top activities and not so much UVBF, therefore UVBF cannot be accountable for the further increase in PTCC post-UVBF, which is most likely due to ongoing generalisation. As PTCC scores had increased post-VAM, the therapy target for block two should not have been velars at all and should have started with /t/. Should this study be carried out again, more careful consideration would be taken in decision making regarding therapy targets and planning. It would also have been useful to have face-to-face inter-rater reliability scoring for interactive discussion regarding complex errors and transcriptions, as suggested by Amorosa et al. (1985). Intra-rater reliability measures showed mostly above average agreement, when there was a three-year gap between transcriptions where the tSLT had gained a vast amount of experience with a range of complex SSDs and UVBF therapy. While the three transcribers all had experience in transcribing complex speech, the use of CLP specialists may have been more suitable for this study.

The study design meant that quantitative ultrasound data was not analysed until after therapy had ceased, with no ultrasound analysis pre-VAM and only qualitative analysis pre-UVBF. As ultrasound data provided crucial diagnostic information which would have led to more accurate knowledge of errors, this may have changed the therapy targets and planning, in turn preventing the errors in clinical decision making. For Craig, therapy block two would have started with targeting /t/ and not velars. For Andrew in therapy block two, gestural instruction would have also focused on pulling the tongue root forward as well as raising the tongue tip for /n/, since tongue root retraction was found in measurements for /n/. This highlights the importance of quantitative analysis of ultrasound data, which is a disadvantage of using ultrasound in terms of time constraints within clinical practice, demonstrating the need for automatic tracking and analysis of ultrasound data to allow more efficient, readily available tools for clinicians.

5.7 Summary and Conclusions

In summary, this thesis aimed to test the effectiveness of UVBF and VAMs for the treatment of SSDs associated with CP. It also investigated the use of ultrasound as a diagnostic tool for the speech of individuals with CP. An overall increase in PTCC scores suggests that combined, both tools were effective in treating the therapy targets for two speakers with repaired submucous cleft palate. Although PTCC scores derived from phonetic transcriptions and listener agreement scores from a multi-listener perceptual evaluation generally corroborate and suggest that in the case of Andrew and Craig VAMs were more effective, due to limitations in the design of the study, it is not possible to conclude that either tool was more effective. Whilst results may suggest VAMs were more effective, both children reported that they preferred UVBF because they were able to see their own tongue moving in real-time.

Variable inter-rater reliability in phonetic transcriptions and the perceptual evaluation confirms some of the issues with perceptual assessment in complex SSDs associated with CP. Articulatory analysis supplemented the perceptual assessment by providing additional information to the phonetic transcriptions, by identifying covert error, with similar findings to previous EPG literature and Bressmann et al. (2011). Interestingly, ultrasound analysis also identified covert contrast in Andrew's data of /n/-/ŋ/ minimal pairs, with significant differences between tongue curves identified in the dorsal and tongue root regions. Surprisingly, Andrew's velar nasal production was indeed produced in a more anterior region than his realisation of /n/, which was retracted further than velar placement, most likely to the uvular region. These subtle differences would not have been detected without quantitative analysis of the articulatory data, suggesting that for diagnostic purposes, high quality recordings and quantitative measurements are essential.

To conclude, both tools have shown promise in the treatment of SSDs associated with CP, however the limitations of the study design highlight the need for future studies to include more rigorous therapeutic designs with a larger number of children with a wider range in Cleft types. With the use of iPads and screen time becoming increasingly popular, the use of commercially available VAMS, such as those in Speech Trainer 3D should be considered for clinical practice, with both children showing more improvement in therapy outcomes with these tools rather than UVBF. As the VAMs in Speech Trainer 3D were not anatomically accurate, it would also be beneficial to test the effectiveness of more accurate models, such as those in Seeing Speech (Lawson et al. 2015).

6 References

ADLER-BOCK, M., BERNHARDT, B., GICK, B. and BACSFALVI, P., 2007. The use of ultrasound in remediation of North American English vertical bar r vertical bar in 2 adolescents. *American Journal of Speech-Language Pathology*. May, vol. 16, no. 2, pp. 128-139.

ALBERT, S., 2005. Einsatz des SpeechTrainers in der Artikulationstherpie bei Kindern. Diplomarbeit, *Studiengang Lehr- und Forschungslogopädie*, RWTH Aachen.

ALEXANDER, K., 2015. *Effectiveness of ultrasound visual biofeedback therapy*. Unpublished Honours Project, Queen Margaret University, Edinburgh.

ALTSHULER, M.W., 1961. A therapeutic oral device for lateral emission. *Journal of Speech and Hearing Disorders*. vol. *26,* no. 2, pp. 179-181.

AMERICAN SPEECH AND HEARING ASSOCIATION (ASHA)., 2017. *Applications (Apps) for Speech-Language Pathology Practice*. [online] [viewed 01 August 2017]. Available from: http://www.asha.org/SLP/schools/Applications-for-Speech-Language-Pathology-Practice/.

AMOROSA, H., VON BENDA, U., WAGNER, E. and KECK, A., 1985. Transcribing phonetic detail in the speech of unintelligible children: A comparison of procedures. *International Journal of Language & Communication Disorders*. Dec, vol. 20, no. 3, pp. 281-287.

APPLE., 2012. *The new iPad*. [online] [viewed 16 March 2012]. Available from: http://store.apple.com/uk/browse/home/shop_ipad/family/ipad.

ARTICULATE INSTRUMENTS LTD., 2010. *Articulate Assistant User Guide: Version 1.18*. Edinburgh, UK: Articulate Instruments Ltd.

ARTICULATE INSTRUMENTS LTD., 2012. *Articulate Assistant Advanced User Guide: Version 2.14*. Edinburgh, UK: Articulate Instruments Ltd.

ARTICULATE INSTRUMENTS LTD., 2015. *Articulate Assistant Advanced User Guide: Version 2.16*. Edinburgh, UK: Articulate Instruments Ltd.

ATKINSON, M. and HOWARD, S., 2011. Physical Structure and Function and Speech Production Associated with Cleft Palate. In: S. HOWARD and A. LOHMANDER, eds. *Cleft palate speech: assessment and intervention*. Chichester: Wiley-Blackwell, pp. 5-22.

AUER, E.T. and BERNSTEIN, L.E., 1997. Speechreading and the structure of the lexicon: computationally modeling the effects of reduced phonetic distinctiveness on lexical uniqueness. *The Journal of the Acoustical Society of America.* Dec, vol. 102, no. 6, pp. 3704-3710.

BACKMAN, C.L. and HARRIS, S.R., 1999. Case studies, single-subject research, and N of 1 randomized trials: comparisons and contrasts. *American Journal of Physical Medicine & Rehabilitation.* Mar-Apr, vol. 78, no. 2, pp. 170-176.

BACKMAN, C.L., HARRIS, S.R., CHISHOLM, J.A. and MONETTE, A.D., 1997. Single-subject research in rehabilitation: a review of studies using AB, withdrawal, multiple baseline, and alternating treatments designs. *Archives of Physical Medicine and Rehabilitation.* Oct, vol. 78, no. 10, pp. 1145-1153.

BACSFALVI, P., 2010. Attaining the lingual components of /r/ with ultrasound for three adolescents with cochlear implants/ Etablissement des composantes linguales du son /r/ a l'aide d'ultrasons chez trois adolescents avec un implant cochleaire. *Canadian Journal of Speech-Language Pathology & Audiology.* Sep, vol. 34, no. 3, pp. 206.

BACSFALVI, P. and BERNHARDT, B.M., 2011. Long-term outcomes of speech therapy for seven adolescents with visual feedback technologies: Ultrasound and electropalatography. *Clinical Linguistics & Phonetics.* Nov/Dec, vol. 25, no. 11/12, pp. 1034.

BACSFALVI, P., BERNHARDT, B.M. and GICK, B., 2007. Electropalatography and ultrasound in vowel remediation for adolescents with hearing impairment. *Advances in Speech Language Pathology.* Jan, vol. 9, no. 1, pp. 36-45.

BADIN, P. and SERRURIER, A., 2006. Three-dimensional linear modelling of tongue: Articulatory data and models. *7th International Seminar on Speech Production, ISSP7.* Ubatuba, Brazil, pp.395-402.

BADIN, P., TARABALKA, Y., ELISEI, F. and BAILLY, G., 2010. Can you ‘read’ tongue movements? Evaluation of the contribution of tongue display to speech understanding. *Speech Communication.* vol. 52, pp. 493-503.

BAKER, E., 2010. Minimal Pair Intervention. In: A.L. WILLIAMS, S. MCLEOD and R.J. MCCAULEY, eds. *Interventions for speech sound disorders in children.* Baltimore, Md.: Paul H. Brookes Pub, pp. 41-72.

BALLARD, K.J., ROBIN, D.A., MCCABE, P. and MCDONALD, J., 2010. A treatment for dysprosody in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research: JSLHR.* Oct, vol. 53, no. 5, pp. 1227-1245.

BARLOW, J.A. and GIERUT, J.A., 2002. Minimal pair approaches to phonological remediation. *Seminars in Speech and Language.* Feb, vol. 23, no. 1, pp. 57-68.

BATES, S. and WATSON, J., 2012. *PPSA (Phonetic and Phonological Systems Analysis)* [online] [viewed 25 September 2017]. Available from: https://www.qmu.ac.uk/schools-and-divisions/shs/ppsa/.

BAYLIS, A., CHAPMAN, K., WHITEHILL, T.L. and GROUP, T.A.S., 2015. Validity and Reliability of Visual Analog Scaling for Assessment of Hypernasality and Audible Nasal Emission in Children With Repaired Cleft Palate. *The Cleft Palate-Craniofacial Journal.* Nov, vol. 52, no. 6, pp. 660-670.

BELLIS, T.H. and WOHLGEMUTH, B., 1999. The incidence of cleft lip and palate deformities in the south-east of Scotland (1971-1990). *British Journal of Orthodontics.* Jun, vol. 26, no. 2, pp. 121-125.

BENOIT, C., LE GOFF, B., 1998. Audio-visual speech synthesis from French test: eight years of models, designs and evaluation at the ICP. *Speech Communication*, vol. 26, pp. 117-129.

BERGAN, C., 2010. Motor Learning Principles and Voice Pedagogy: Theory and Practice. *Journal of Singing.* /03/01, vol. 66, no. 4, pp. 457.

BERNHARDT, B., GICK, B., BACSFALVI, P. and ADLER-BOCK, M., 2005. Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics & Phonetics.* Sep-Nov, vol. 19, no. 6-7, pp. 605-617.

BERNHARDT, B., GICK, B., BACSFALVI, P. and ASHDOWN, J., 2003. Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics & Phonetics.* Jan, vol. 17, no. 3, pp. 199-216.

BERNHARDT, M.B., BACSFALVI, P., ADLER-BOCK, M., SHIMIZU, R., CHENEY, A., GIESBRECHT, N., O'CONNELL, M., SIRIANNI, J. and RADANOV, B., 2008. Ultrasound as visual feedback in speech habilitation: Exploring consultative use in rural British Columbia, Canada. *Clinical Linguistics & Phonetics.* Jan, vol. 22, no. 2, pp. 149-162.

BESSELL, A., SELL, D., WHITING, P., ROULSTONE, S., ALBERY, L., PERSSON, M., VERHOEVEN, A., BURKE, M. and NESS, A.R., 2013. Speech and language therapy interventions for children with cleft palate: a systematic review. *The Cleft Palate-Craniofacial Journal.* Jan, vol. 50, no. 1, pp. e1-e17.

BLYTH, K.M., MCCABE, P., MADILL, C. and BALLARD, K.J., 2016. Ultrasound visual feedback in articulation therapy following partial glossectomy. *Journal of Communication Disorders.* May-Jun, vol. 61, pp. 1-15.

BOERSMA, P. and WEENINK, D., 2013. *PRAAT doing phonetics by computer*. Version 5.3.57. [online] [viewed 10 March 2014]. Available from: www.praat.org.

BRESSMANN, T., FLOWERS, H., WONG, W. and IRISH, J.C., 2010. Coronal view ultrasound imaging of movement in different segments of the tongue during paced recital: findings from four normal speakers and a speaker with partial glossectomy. *Clinical Linguistics & Phonetics.* Aug, vol. 24, no. 8, pp. 589-601.

BRESSMANN, T., HARPER, S., ZHYLICH, I. and KULKARNI, G.V., 2016. Perceptual, durational and tongue displacement measures following articulation therapy for rhotic sound errors. *Clinical Linguistics & Phonetics.* vol. 30, no. 3-5, pp. 345-362.

BRESSMANN, T., RADOVANOVIC, B., KULKARNI, G.V., KLAIMAN, P. and FISHER, D., 2011. An ultrasonographic investigation of cleft-type compensatory articulations of voiceless velar stops. *Clinical Linguistics & Phonetics.* Nov, vol. 25, no. 11-12, pp. 1028-1033.

BRITTON, L., ALBERY, L., BOWDEN, M., HARDING-BELL, A., PHIPPEN, G. and SELL, D., 2014. A Cross-Sectional Cohort Study of Speech in Five-Year-Olds With Cleft Palate +/- Lip to Support Development of National Audit Standards BENCHMARKING SPEECH STANDARDS IN THE UNITED KINGDOM. *Cleft Palate-Craniofacial Journal.* Jul, vol. 51, no. 4, pp. 431-451.

BROBECK, T.C. and LUBINSKY, J., 2003. Using Single-Subject Designs in Speech-Language Pathology Practicum. *Contemporary Issues in Communication Science and Disorders*. vol. 30, pp. 101-106.

BRUNNEGÅRD, K. and LOHMANDER, A., 2007. A cross-sectional study of speech in 10-year-old children with cleft palate: results and issues of rater reliability. *The Cleft Palate-Craniofacial Journal.* Jan, vol. 44, no. 1, pp. 33-44.

BRUNNEGÅRD, K., LOHMANDER, A. and DOORN, J.V., 2009. Untrained listeners' ratings of speech disorders in a group with cleft palate: a comparison with speech and language pathologists, ratings. *International Journal of Language & Communication Disorders.* Jan, vol. 44, no. 5, pp. 656-674.

BUCKINGHAM, H.W. and YULE, G. 1987. Phonemic false evaluation: Theoretical and clinical aspects. *Clinical Linguistics & Phonetics.* Jan, vol. 1, no. 2, pp. 113-125.

BUGAIGHIS, I., O'HIGGINS, P., TIDDEMAN, B., MATTICK, C., BEN ALI, O. and HOBSON, R., 2010. Three-dimensional geometric morphometrics applied to the study of children with cleft lip and/or palate from the North East of England. *European Journal of Orthodontics.* Oct, vol. 32, no. 5, pp. 514-521.

CAOUETTE-LABERGE, L., BAYET, B. and LAROCQUE, Y., 1994. The Pierre Robin Sequence: Review of 125 Cases and Evolution of Treatment Modalities. *Plastic and Reconstructive Surgery.* Apr, vol. 93, no. 5, pp. 934.

CARTER, P. and EDWARDS, S., 2004. EPG therapy for children with long-standing speech disorders: predictions and outcomes. *Clinical Linguistics & Phonetics.* Sep-Dec, vol. 18, no. 6-8, pp. 359-372.

CARUSO, A.J. and STRAND, E.A., 1999. Motor speech disorders in children: Definitions, backgrounds and a theoretical framework. In A. J. CARUSO and E. A. STRAND, eds. *Clinical management of motor speech disorders in children*. New York: Thieme, pp.1-27.

CASTICK, S., KNIGHT, R. and SELL, D., 2017. Perceptual Judgments of Resonance, Nasal Airflow, Understandability, and Acceptability in Speakers With Cleft Palate: Ordinal Versus Visual Analogue Scaling. *The Cleft Palate-Craniofacial Journal.* Jan, vol. 54, no. 1, pp. 19-31.

CHAPMAN, K.L., 1993. Phonologic processes in children with cleft palate. *The Cleft Palate-Craniofacial Journal.* Jan, vol. 30, no. 1, pp. 64-72.

CHAPMAN, K.L. and HARDIN, M.A., 1992. Phonetic and phonologic skills of two-year-olds with cleft palate. *The Cleft Palate-Craniofacial Journal.* Sep, vol. 29, no. 5, pp. 435-443.

CHAPMAN, K.L., HARDIN-JONES, M. and HALTER, K.A., 2003. The relationship between early speech and later speech and language performance for children with cleft lip and palate. *Clinical Linguistics & Phonetics.* Jan, vol. 17, no. 3, pp. 173-197.

CHAPMAN, K.L., HARDIN-JONES, M.A., GOLDSTEIN, J.A., HALTER, K.A., HAVLIK, R.J. and SCHULTE, J., 2008. Timing of palatal surgery and speech outcome. *The Cleft Palate-Craniofacial Journal.* May, vol. 45, no. 3, pp. 297-308.

CHAPMAN, K.L., HARDIN-JONES, M., SCHULTE, J. and HALTER, K.A., 2001. Vocal development of 9-month-old babies with cleft palate. *Journal of Speech, Language, and Hearing Research: JSLHR.* Dec, vol. 44, no. 6, pp. 1268-1283.

CHAPMAN, K.L. and WILLADSEN, E., 2011. The Development of Speech in Children with Cleft Palate. In: S. HOWARD and A. LOHMANDER, eds. *Cleft palate speech: assessment and intervention.* Chichester: Wiley-Blackwell, pp. 23-40.

CLEFT CARE SCOTLAND., 2016. *Cleft Care Scotland National Managed Clinical Network Annual Report 2015/16*. [online] [viewed 01 August 2017]. Available from: http://www.mcns.scot.nhs.uk/types-of-network/national-networks-in-scotland/nmcns/ccs/.

CLELAND, J., 2017-2018. *Visualising Speech: Using Ultrasound Visual Biofeedback to Diagnose and Treat Speech Disorders in Children with Cleft Lip and Palate*. [online] [viewed 01 October 2017]. Available from: https://pure.strath.ac.uk/portal/en/projects/visualising-speech-using-ultrasound-visual-biofeedback-to-diagnose-and-treat-speech-disorders-in-children-with-cleft-lip-and-palate(50210cac-51c0-4fb7-b73e-77b3aedbe1b9).html.

CLELAND, J., CRAMPIN, L., WRENCH, A.A., ZHARKOVA N. and LLOYD, S., 2017a. Royal College of Speech and Language Therapists Conference 2017. *Visualising speech: using ultrasound to diagnose and treat speech disorders in children with cleft lip and palate* [poster]. Glasgow, United Kingdom. 27-28 September.

CLELAND, J. and ISLES, J., 2017. *Evidence of efficacy: Ultrasound speech therapy studies (July 2017).* [online] [viewed 10 August 2017]. Available from: http://www.articulateinstruments.com/US_Treat_Refs_0717.pdf.

CLELAND, J., MCCRON, C. and SCOBBIE, J.M., 2013. Tongue reading: Comparing the interpretation of visual information from inside the mouth, from electropalatographic and ultrasound displays of speech sounds. *Clinical Linguistics & Phonetics.* Apr, vol. 27, no. 4, pp. 299-311.

CLELAND, J. and SCOBBIE, J.M., In Press. Acquisition of New Speech Motor Plans Via Articulatory Visual Biofeedback. In S. FUCHS, J. CLELAND, and A. ROCHET-CAPPELAN, eds. *Speech Perception and Production: Learning and Memory*. Berlin, Germany: Peter Lang.

CLELAND, J., SCOBBIE, J.M., HEYDE, C.J., ROXBURGH, Z. and WRENCH, A.A., 2017b. Covert contrast and covert errors in persistent velar fronting. *Clinical Linguistics & Phonetics.* vol. 31, no. 1, pp. 35-55.

CLELAND, J., SCOBBIE, J.M., ISLES, J. and ALEXANDER, K., 2015a. Ultrafest VII. *Gradient acquisition of velars via Ultrasound Visual Biofeedback for persistent velar fronting* [poster]. Hong Kong: The University of Hong Kong, December.

CLELAND, J., SCOBBIE, J.M., NAKAI, S. and WRENCH, A.A., 2015b. Helping children learn non-native articulations: the implications for ultrasound-based clinical intervention. In: THE SCOTTISH CONSORTIUM FOR ICPHS 2015, ed. *Proceedings of the 18th International Congress of Phonetic Sciences: ICPhS 2015*. Scotland, pp. 1-5.

CLELAND, J., SCOBBIE, J.M. and WRENCH, A.A., 2015c. Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics & Phonetics.* vol. 29, no. 8-10, pp. 575-597.

CLELAND, J., SCOBBIE, J.M., ROXBURGH, Z., HEYDE, C. and WRENCH, A.A., 2017c. 7th International Conference on Speech Motor Control. *Ultraphonix: using ultrasound visual biofeedback to teach children with special speech sound disorders new articulations* [poster]. Groningen, Netherlands, July.

CLELAND, J., TIMMINS, C., WOOD, S.E., HARDCASTLE, W.J. and WISHART, J.G., 2009. Electropalatographic therapy for children and young people with Down's syndrome. *Clinical Linguistics & Phonetics.* Dec, vol. 23, no. 12, pp. 926-939.

COHEN, J., 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement.* Apr, vol. 20, no. 1, pp. 37-46.

COHEN, M.M., 1978. Syndromes with cleft lip and cleft palate. *The Cleft Palate Journal.* Oct, vol. 15, no. 4, pp. 306-328.

COLEMAN, J.R. and SYKES, J.M., 2001. The embryology, classification, epidemiology, and genetics of facial clefting. *Facial Plastic Surgery Clinics of North America.* /02, vol. 9, no. 1, pp. 1-13.

COMMON SENSE MEDIA., 2013. *Zero to Eight: Children's Media Use in America 20*13. [online] [viewed 01 January 2015]. Available from: www.commonsensemedia.org.

CORDER, G.W. and FOREMAN, D.I., 2014. *Nonparametric Statistics: A Step-by-Step Approach.* 2nd ed. Hoboken, New Jersey: Wiley.

CORDES, A.K., 1994. The reliability of observational data: I. Theories and methods for speech-language pathology. *Journal of Speech and Hearing Research.* Apr, vol. 37, no. 2, pp. 264-278.

CROSS, E.S., HAMILTON, A.F.D.C. and GRAFTON, S.T., 2006. Building a motor simulation de novo: observation of dance by dancers. *NeuroImage.* Jul 01, vol. 31, no. 3, pp. 1257-1267.

DALSTON, R.M., MARSH, J.L., VIG, K.W., WITZEL, M.A. and BUMSTED, R.M., 1988. Minimal standards for reporting the results of surgery on patients with cleft lip, cleft palate, or both: a proposal. *The Cleft Palate Journal.* Jan, vol. 25, no. 1, pp. 3-7.

DAVIDSON, L., 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America.* Jul, vol. 120, no. 1, pp. 407-415.

DIXON, M.J., MARAZITA, M.L., BEATY, T.H. and MURRAY, J.C., 2011. Cleft lip and palate: understanding genetic and environmental influences. *Nature Reviews. Genetics.* Mar, vol. 12, no. 3, pp. 167-178.

DODD, B., 2005. *Differential diagnosis and treatment of children with speech disorder.* London: Whurr.

DODD, B., HUA, Z., CROSBIE, S., HOLM, A. and OZANNE, A., 2002. *Diagnostic Evaluation of Articulation and Phonology.* London: The Psychological Corporation.

DUNHAM, G., 2011. The Future at Hand: Mobile Devices and Apps in Clinical Practice. *The ASHA Leader.* vol. 16, no. 4, pp. 4-4.

DUNN, L.M., DUNN, D.M. and NATIONAL FOUNDATION FOR EDUCATIONAL RESEARCH., 2009. *British Picture Vocabulary Scale 3rd ed. (BPVSIII).* Great Britain: Wascana Ltd partnership and GL Assessment.

FAGEL, S. and MADANY, K., 2008. A 3-D virtual head as a tool for speech therapy for children. *Interspeech 2008*. Brisbane, Australia, pp. 2643–2646.

FAN, W.S., MULLIKEN, J.B. and PADWA, B.L., 2005. An association between hemifacial microsomia and facial clefting. *Journal of Oral and Maxillofacial Surgery.* Mar, vol. 63, no. 3, pp. 330-334.

FAWCETT, S., BACSFALVI, P. and BERNHARDT, B. M., 2008. Ultrasound as visual feedback in speech therapy for /r/ with adults with Down Syndrome. *Down Syndrome Quarterly*. vol. 10, no. 1, pp. 4–12.

FERNANDES, B., 2011. iTherapy: The Revolution of Mobile Devices Within the Field of Speech Therapy. *Perspectives on School-Based Issues.* Jun, vol. 12, no. 2, pp. 35-40.

FLEISS, J. L., 1981. *Statistical methods for rates and proportions*. New York: Wiley.

FLETCHER, S.G., 1972. Contingencies for Bioelectronic Modification of Nasality. *Journal of Speech and Hearing Disorders.* Aug, vol. 37, no. 3, pp. 329-346.

FORREST, K., 2002. Are oral-motor exercises useful in the treatment of phonological/articulatory disorders? *Seminars in Speech and Language.* Feb, vol. 23, no. 1, pp. 15-26.

FRANCE, R.B. and DEANGELO, L., 2016. Biofeedback. *Magill's Medical Guide.* [online] [viewed 15 August 2017]. Available from: http://eds.a.ebscohost.com/eds/detail/detail?vid=6&sid=9e7a7908-14e2-4f20-822d-f9d6bb529002%40sessionmgr4006&bdata=JnNpdGU9ZWRzLWxpdmU%3d#AN=87690452&db=ers.

FRIEL, S., 1998. When is a /k/ not a [k]? EPG as a diagnostic and therapeutic tool for abnormal velar stops. *International Journal of Language & Communication Disorders.* vol. 33, pp. 439-444.

GARCÍA VELASCO, M., YSUNZA, A., HERNANDEZ, X. and MARQUEZ, C., 1988. Diagnosis and treatment of submucous cleft palate: a review of 108 cases. *The Cleft Palate Journal.* Apr, vol. 25, no. 2, pp. 171-173.

GIBBON, F., 1990. Lingual activity in two speech-disordered children's attempts to produce velar and alveolar stop consonants: evidence from electropalatographic

(EPG) data. *The British Journal of Disorders of Communication.* Dec, vol. 25, no. 3, pp. 329-340.

GIBBON, F.E., 2004. Abnormal patterns of tongue-palate contact in the speech of individuals with cleft palate. *Clinical Linguistics & Phonetics.* Jun, vol. 18, no. 4-5, pp. 285-311.

GIBBON, F.E., 1999. Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research: JSLHR.* Apr, vol. 42, no. 2, pp. 382-397.

GIBBON, F.E. and CRAMPIN, L., 2001. An electropalatographic investigation of middorsum palatal stops in an adult with repaired cleft palate. *The Cleft Palate-Craniofacial Journal.* Mar, vol. 38, no. 2, pp. 96-105.

GIBBON, F., HARDCASTLE, W.J., CRAMPIN, L., REYNOLDS, B., RAZZELL, R. and WILSON, J., 2001. Visual feedback therapy using electropalatography (EPG) for articulation disorders associated with cleft palate. *Asia Pacific Journal of Speech, Language and Hearing.* Jan, vol. 6, no. 1, pp. 53-58.

GIBBON, F., HARDCASTLE, B. and DENT, H., 1995. A study of obstruent sounds in school-age children with speech disorders using electropalatography. *European Journal of Disorders of Communication.* vol. 30, no. 2, pp. 213-225.

GIBBON, F.E., ELLIS, L. and CRAMPIN, L., 2004. Articulatory placement for t, d, k and g targets in school age children with speech disorders associated with cleft palate. *Clinical Linguistics & Phonetics.* Sep, vol. 18, no. 6-8, pp. 391-404.

GIBBON, F.E. and LEE, A., 2011. Articulation – Instruments for Research and Clinical Practice. In: S. HOWARD and A. LOHMANDER, eds. *Cleft palate speech: assessment and intervention.* Chichester: Wiley-Blackwell, pp. 221-238.

GIBBON, F.E. and LEE, A., 2017. Electropalatographic (EPG) evidence of covert contrasts in disordered speech. *Clinical Linguistics & Phonetics.* Jan 2, vol. 31, no. 1, pp. 4-20.

GIBBON, F.E., YUEN, I., LEE, A. and ADAMS, L., 2007. Normal adult speakers' tongue palate contact patterns for alveolar oral and nasal stops. *International Journal of Speech-Language Pathology.* vol. 9, no. 1, pp. 82-89.

GIBBON, F., CRAMPIN, L., HARDCASTLE, B., NAIRN, M., RAZZELL, R., HARVEY, L. and REYNOLDS, B., 1998. CLEFTNET (Scotland): A Network for the Treatment of Cleft Palate Speech Using Epg. *International Journal of Language & Communication Disorders.* Dec, vol. 33, pp. 44.

GIBBON, F. and SCOBBIE, J. M., 1997. Covert contrasts in children with phonological disorder. *Australian Communication Quarterly.*, Autumn, pp. 13-16.

GIBBON, F.E. and WOLTERS, M., 2005. Craniofacial Society of Great Britain and Ireland Annual Scientific Conference. *A new application of ultrasound to image tongue behaviour in cleft palate speech* [Poster]. Swansea, UK, 13-15 April.

GIBBON, F.E. and WOOD, S. E., 2010. Visual feedback Therapy with Electropalatography. In: A.L. WILLIAMS, S. MCLEOD and R.J. MCCAULEY, eds. *Interventions for speech sound disorders in children.* Baltimore, Md.: Paul H. Brookes Pub, pp. 509-536.

GOLDING-KUSHNER, K.J., 1995. Treatment of articulation and resonance disorders associated with cleft palate management and VPI. In: R. SHPRINTZEN and J. BARDACH, eds. *Cleft Palate Speech Management: A Multidisciplinary approach*. St Louis: Mosby, pp. 327-351.

GOLDING-KUSHNER, K.J., WELLER, G. and SHPRINTZEN, R.J., 1985. Velo-cardio-facial syndrome: language and psychological profiles. *Journal of Craniofacial Genetics and Developmental Biology.* vol. 5, no. 3, pp. 259-266.

GOOCH, J.L., HARDIN-JONES, M., CHAPMAN, K.L., TROST-CARDAMONE, J.E. and SUSSMAN, J., 2001. Reliability of listener transcriptions of compensatory articulations. *The Cleft Palate-Craniofacial Journal.* Jan, vol. 38, no. 1, pp. 59-67.

GOSAIN, A.K., CONLEY, S.F., MARKS, S. and LARSON, D.L., 1996. Submucous Cleft Palate: Diagnostic Methods and Outcomes of Surgical Treatment. *Plastic and Reconstructive Surgery.* Jun, vol. 97, no. 7, pp. 1497.

GOSNELL, J., 2011. Apps: An Emerging Tool for SLPs: A plethora of apps can be used to develop expressive, receptive, and other language skills. *The ASHA Leader.* vol. 16, no. 12, pp. 10-13.

GOTTO, J., 2004. Therapie der Sprechapraxie: Eine Einzelfallstudie zum PC-Programm SpeechTrainer. *Diplomarbeit, Studiengang Lehr- und Forschungslogopädie*, RWTH Aachen.

GUADAGNOLI, M.A. and LEE, T.D., 2004. Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior.* Jun, vol. 36, no. 2, pp. 212-224.

HARDCASTLE, W.J. and GIBBON, F., 2005. Electropalatography as a research and clinical tool. 30 Years on. In W.J. HARDCASTLE and J.M. BECK, eds. *A Figure of Speech: A Festschrift for John Laver.* New York: Routledge, pp. 39–62.

HARDCASTLE, W., MORGAN, B.R. and NUNN, M., 1989, Instrumental articulatory phonetics in assessment and remediation: case studies with the electropalatograph. In J. STENGELHOFEN, ed. *Cleft Palate: The Nature and Remediation of Communication Problems*. Edinburgh: Churchill Livingstone, pp. 136–164.

HARDCASTLE, W.J. and MORGAN, R.A., 1982. An instrumental investigation of articulation disorders in children. *The British Journal of Disorders of Communication.* Apr, vol. 17, no. 1, pp. 47-65.

HARDCASTLE, W.J. and LAVER, J., 1997. *Handbook of phonetic sciences.* Oxford: Blackwell.

HARDING, A. and GRUNWELL, P., 1993. The relationship between speech and timing of hard palate repair. In: P. GRUNWELL, ed. *Analysing Cleft Palate Speech*. London: Whurr Publishers, pp. 48-82.

HARDING, A. and GRUNWELL, P., 1996. Characteristics of cleft palate speech. *European Journal of Disorders of Communication: The Journal of the College of Speech and Language Therapists, London.* vol. 31, no. 4, pp. 331-357.

HARDING, A. and GRUNWELL, P., 1998. Active versus passive cleft-type speech characteristics. *International Journal of Language & Communication Disorders.* Jul, vol. 33, no. 3, pp. 329-352.

HARDING-BELL, A. and HOWARD, A., 2011. Phonological Approaches to Speech Difficulties Associated with Cleft Palate. In: S. HOWARD and A. LOHMANDER, eds. *Cleft palate speech: assessment and intervention.* Chichester: Wiley-Blackwell, pp. 275-291.

HARDIN-JONES, M.A. and JONES, D.L., 2005. Speech production of preschoolers with cleft palate. *The Cleft Palate-Craniofacial Journal.* Jan, vol. 42, no. 1, pp. 7-13.

HARDIN-JONES, M., CHAPMAN, K.L. and SCHULTE, J., 2003. The impact of cleft type on early vocal development in babies with cleft palate. *The Cleft Palate-Craniofacial Journal.* Sep, vol. 40, no. 5, pp. 453-459.

HARVILLE, E.W., WILCOX, A.J., LIE, R.T., VINDENES, H. and ABYHOLM, F., 2005. Cleft lip and palate versus cleft lip only: are they distinct defects? *American Journal of Epidemiology.* Sep, vol. 162, no. 5, pp. 448-453.

HENG, Q., MCCABE, P., CLARKE, J. and PRESTON, J.L., 2016. Using ultrasound visual feedback to remediate velar fronting in preschool children: A pilot study. *Clinical Linguistics & Phonetics.* vol. 30, no. 3-5, pp. 382-397.

HESELWOOD, B. and HOWARD, S., 2008. Clinical Phonetic Transcription. In: M.J. BALL, M.R. PERKINS, N. MULLER and S. HOWARD, eds. *The Handbook of Clinical Linguistics*. USA, UK, Australia: Blackwell Publishing, pp. 381-399.

HEWLETT, N., 1988. Acoustic properties of /k/ and /t/ in normal and phonologically disordered speech. *Clinical Linguistics & Phonetics.* Jan, vol. 2, no. 1, pp. 29-45.

HEWLETT, N. and BECK, J.M., 2006. *An introduction to the science of phonetics.* Lawrence Erlbaum Associates.

HITCHCOCK, E.R. and MCALLISTER BYUN, T., 2015. Enhancing generalisation in biofeedback intervention using the challenge point framework: A case study. *Clinical Linguistics & Phonetics.* Jan, vol. 29, no. 1, pp. 59-75.

HODSON, B., 2010. *Evaluating and enhancing children's phonological systems: Research and theory to practice.* Wichita, KS: PhonoComp.

HODSON, B.W., CHIN, L., REDMOND, B. and SIMPSON, R., 1983. Phonological evaluation and remediation of speech deviations of a child with a repaired cleft palate: a case study. *The Journal of Speech and Hearing Disorders.* Feb, vol. 48, no. 1, pp. 93-98.

HOWARD, S., 1998, A perceptual and electropalatographic case study of Pierre Robin Sequence. In W. ZIEGLER and K. DEGER, eds. *Clinical Phonetics and Linguistics*. London: Whurr Publishers, pp. 157–164.

HOWARD, S., 2004. Compensatory articulatory behaviours in adolescents with cleft palate: comparing the perceptual and instrumental evidence. *Clinical Linguistics & Phonetics.* Jun, vol. 18, no. 4-5, pp. 313-340.

HOWARD, S., 2011. Phonetic Transcription for Speech Related to Cleft Palate. In: S. HOWARD and A. LOHMANDER, eds. *Cleft palate speech: assessment and intervention.* Chichester: Wiley-Blackwell, pp. 127-144.

HOWARD, S. and PICKSTONE, C., 1995, Cleft palate: perceptual and instrumental analysis of a phonological system. In M. PERKINS and S. HOWARD, eds. *Case Studies in Clinical Linguistics*. London: Whurr Publishers, pp. 65–90.

INTERNATIONAL PHONETICS ASSOCIATION (IPA)., 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

IRWIN, D., PANNBACKER, M.H. and LASS, N.J., 2013. *Clinical research methods in speech-language pathology and audiology, 2nd ed.* San Diego: Plural Publishing Inc.

ISOTALO, E., PULKKINEN, J. and HAAPANEN, M., 2007. Speech in 6 year old children with sub-mucous cleft palate. *The Journal of Craniofacial Surgery.* Jul, vol. 18, no. 4, pp. 724.

JOFFE, V. and PRING, T., 2008. Children with phonological problems: a survey of clinical practice. *International Journal of Language & Communication Disorders.* Mar-Apr, vol. 43, no. 2, pp. 154-164.

JOHN, A., SELL, D., HARDING-BELL, A., SWEENEY, T. and WILLIAMS, A., 2003. The development of a valid and reliable tool for auditing speech outcome in cleft care. [Conference Presentation]. *Craniofacial Society of Great Britain and Ireland*. Leeds, UK.

JONES, C.E., CHAPMAN, K.L. and HARDIN-JONES, M.A., 2003. Speech development of children with cleft palate before and after palatal surgery. *The Cleft Palate-Craniofacial Journal.* Jan, vol. 40, no. 1, pp. 19-31.

KAPLAN, E.N., 1975. The occult submucous cleft palate. *The Cleft Palate Journal.* Oct, vol. 12, pp. 356-368.

KATZ, W.F., MCNEIL, M.R. and GARST, D.M., 2010. Treating apraxia of speech (AOS) with EMA-supplied visual augmented feedback. *Aphasiology.* Jul, vol. 24, no. 6-8, pp. 826-837.

KENT, R.D., WEISMER, G., KENT, J.F., VORPERIAN, H.K. and DUFFY, J.R., 1999. Acoustic studies of dysarthric speech: methods, progress, and potential. *Journal of Communication Disorders.* May-Jun, vol. 32, no. 3, pp. 189.

KEUNING, K.H., WIENEKE, G.H. and DEJONCKERE, P.H., 1999. The intrajudge reliability of the perceptual rating of cleft palate speech before and after pharyngeal flap surgery: the effect of judges and speech samples. *The Cleft Palate-Craniofacial Journal.* Jul, vol. 36, no. 4, pp. 328-333.

KILPELÄINEN, P.V. and LAINE-ALAVA, M.T., 1996. Palatal asymmetry in cleft palate subjects. *The Cleft Palate-Craniofacial Journal.* Nov, vol. 33, no. 6, pp. 483-488.

KLEIN, H.B., BYUN, T.M., DAVIDSON, L. and GRIGOS, M.I., 2013. A Multidimensional Investigation of Children's /r/ Productions: Perceptual, Ultrasound, and Acoustic Measures. *American Journal of Speech-Language*. vol. 22, no. 3, pp. 540-553.

KORNFELD, J. R., 1971. Theoretical issues in child phonology. In: CHICAGO LINGUISTIC SOCIETY, ed. *Chicago Linguistic Society 7thth Regional Meeting*, pp. 454–468.

KOSOWSKI, T., WEATHERS, W., WOLFSWINKEL, E. and RIDGWAY, E., 2012. Cleft Palate. *Seminars in Plastic Surgery.* Nov, vol. 26, no. 4, pp. 164-169.

KRÖGER, B.J., 2003. Ein visuelles Modell der Artikulation. *Laryngo-Rhino-Otologie.* vol. 82, no. 06, pp. 402-407.

KRÖGER, B.J., GOTTO, J., ALBERT, S. and NEUSCHAEFER-RUBE, C., 2005. A visual articulatory model and its application to therapy of speech disorders: a pilot study. *ZAS Papers in Linguistics*. vol. 40, pp 79-94.

KRÖGER, B.J., GRAF-BORTTSCHELLER, V. and LOWIT, A., 2008. Two- and Three-Dimensional Visual Articulatory Models for Pronunciation Training and for

Treatment of Speech Disorders. *ISCA*. Brisbane, Australia. September 22-26, pp. 2639-2642.

KUEHN, D.P. and MOLLER, K.T., 2000. Speech and Language Issues in the Cleft Palate Population: The State of the Art. *The Cleft Palate-Craniofacial Journal.* Jun, vol. 37, no. 4, pp. 348.

KUMMER, A.W., 2001. *Cleft palate and craniofacial anomalies: the effects on speech and resonance.* San Diego, Calif: Singular/Thomson Learning.

KUMMER, A.W., 2014. Speech evaluation for patients with cleft palate. *Clinics in Plastic Surgery.* Apr, vol. 41, no. 2, pp. 241-251.

KURT, G., BAYRAM, M., UYSAL, T. and OZER, M., 2010. Mandibular asymmetry in cleft lip and palate patients. *European Journal of Orthodontics.* vol. 32, no. 1, pp. 19-23.

KYLE, F.E., CAMPBELL, R., MOHAMMED, T., COLEMAN, M. and MACSWEENEY, M., 2013. Speechreading development in deaf and hearing children: introducing the test of child speechreading. *Journal of Speech, Language, and Hearing Research: JSLHR.* Apr, vol. 56, no. 2, pp. 416.

LADEFOGED, P. and MADDIESON, I., 1996. *The sounds of the world's languages.* Oxford: Blackwell.

LANDIS, J.R. and KOCH, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics.* vol. 33, no. 1, pp. 159-174.

LAWRENCE, C.W. and PHILIPS, B.J., 1975. A telefluoroscopic study of lingual contacts made by persons with palatal defects. *The Cleft Palate Journal.* Jan, vol. 12, no. 00, pp. 85-94.

LAWSON, E., STUART-SMITH, J., SCOBBIE, J.M. and NAKAI, S., 2015. *Seeing Speech: An Articulatory Web Resource for the Study of Phonetics.* [online] [viewed 01 July 2017]. Available from: http://www.seeingspeech.ac.uk/.

LEE, A., GIBBON, F.E., CRAMPIN, L., YUEN, I. and MCLENNAN, G., 2007. The national CLEFTNET project for individuals with speech disorders associated with cleft palate. *International Journal of Speech-Language Pathology.* vol. 9, no. 1, pp. 57-64.

LEE, A.S., LAW, J. and GIBBON, F.E., 2009. Electropalatography for articulation disorders associated with cleft palate (Review). *The Cochrane Collaboration*. London: John Wiley & Sons, Ltd.

LEE, S.A.S., WRENCH, A.A. and SANCIBRIAN, S., 2015. How To Get Started With Ultrasound Technology for Treatment of Speech Sound Disorders. *Perspectives on Speech Science and Orofacial Disorders.* vol. 25, no. 2, pp. 66.

LEVI, B., BRUGMAN, S., WONG, V.W., GROVA, M., LONGAKER, M.T. and WAN, D.C., 2011. Palatogenesis. *Organogenesis.* Oct, vol. 7, no. 4, pp. 242.

LIBERMAN, A.M., HARRIS, K.S., HOFFMAN, H.S. and GRIFFITH, B.C., 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology.* Nov, vol. 54, no. 5, pp. 358-368.

LIBERMAN, A.M. and MATTINGLY, I.G., 1985. The motor theory of speech perception revised. *Cognition.* Oct, vol. 21, no. 1, pp. 1-36.

LIEBERMAN, P., CRELIN, E.S. and KLATT, D.H., 1972. Phonetic Ability and Related Anatomy of the Newborn and Adult Human, Neanderthal Man, and the Chimpanzee. *American Anthropologist.* vol. 74, no. 3, pp. 287-307.

LIPETZ, H.M. and BERNHARDT, B.M., 2013. A multi-modal approach to intervention for one adolescent's frontal lisp. *Clinical Linguistics & Phonetics.* Jan, vol. 27, no. 1, pp. 1-17.

LISMAN, A.L. and SADAGOPAN, N., 2013. Focus of attention and speech motor performance. *Journal of Communication Disorders.* May-Jun, vol. 46, no. 3, pp. 281-293.

LOHMANDER, A., 2011. Surgical Intervention and Speech Outcomes in Cleft Lip and Palate. In: S. HOWARD and A. LOHMANDER, eds. *Cleft palate speech: assessment and intervention.* Chichester: Wiley-Blackwell, pp. 55-85.

LOHMANDER, A., HENRIKSSON, C. and HAVSTAM, C., 2010. Electropalatography in home training of retracted articulation in a Swedish child with cleft palate: Effect on articulation pattern and speech. *International Journal of Speech-Language Pathology.* Dec, vol. 12, no. 6, pp. 483-496.

LOHMANDER, A. and OLSSON, M., 2004. Methodology for perceptual assessment of speech in patients with cleft palate: a critical review of the literature. *The Cleft Palate-Craniofacial Journal.* vol. 41, no. 1, pp. 64-70.

LOHMANDER, A. and PERSSON, C., 2008. A longitudinal study of speech production in Swedish children with unilateral cleft lip and palate and two-stage palatal repair. *The Cleft Palate-Craniofacial Journal.* Jan, vol. 45, no. 1, pp. 32-41.

LOHMANDER-AGERSKOV, A., SÖDERPALM, E., FRIEDE, H., PERSSON, E.C. and LILJA, J., 1994. Pre-speech in children with cleft lip and palate or cleft palate only: phonetic analysis related to morphologic and functional factors. *The Cleft Palate-Craniofacial.* Jul, vol. 31, no. 4, pp. 271-279.

LUIJSTERBURG, A.J.M. and VERMEIJ-KEERS, C., 2011. Ten years recording common oral clefts with a new descriptive system. *The Cleft Palate-Craniofacial Journal.* Mar, vol. 48, no. 2, pp. 173.

LYNCH, J.I., FOX, D.R. and BROOKSHIRE, B.L., 1983. Phonological proficiency of two cleft palate toddlers with school-age follow-up. *The Journal of Speech and Hearing Disorders.* Aug, vol. 48, no. 3, pp. 274-285.

MAAS, E., BUTALLA, C.E. and FARINELLA, K.A., 2012. Feedback frequency in treatment for childhood apraxia of speech. *American Journal of Speech-Language Pathology.* Aug, vol. 21, no. 3, pp. 239-257.

MAAS, E. and FARINELLA, K.A., 2012. Random versus blocked practice in treatment for childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research: JSLHR.* Apr, vol. 55, no. 2, pp. 561-578.

MAAS, E., ROBIN, D.A., AUSTERMANN HULA, S.N., FREEDMAN, S.E., WULF, G., BALLARD, K.J. and SCHMIDT, R.A., 2008. Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology.* Aug, vol. 17, no. 3, pp. 277-298.

MACKEN, M.A. and BARTON, D., 1980. The acquisition of the voicing contrast in English: study of voice onset time in word-initial stop consonants. *Journal of Child Language.* Feb, vol. 7, no. 1, pp. 41-74.

MARSH, J., PLOWMAN, L., YAMADA-RICE, D., BISHOP, J.C., LAHMAR, J., SCOTT, F., DAVENPORT, A., DAVIS, S., FRENCH, K., PIRAS, M., THORNHILL, S., ROBINSON, P. and WINTER, P., 2015. *Exploring Play and Creativity in Pre-Schoolers' Use of Apps: Final Project Report*. [online] [viewed 01 August 2017]. Available from: www.techandplay.org.

MARSHALL, J., 2001. The sociolinguistic status of the glottal stop in Northeast Scots. *Reading Working Papers in Linguistics 5*, pp. 49–65.

MASSARO, D., BIGLER, S., CHEN, T., PERLMAN, M. and OUNI, S., 2008. Pronunciation Training: The Role of Ear and Eye. *Interspeech 2008*, 22-26 September, Brisbane, Australia.

MCALLISTER BYUN, T., BUCHWALD, A. and MIZOGUCHI, A., 2016. Covert contrast in velar fronting: An acoustic and ultrasound study. *Clinical Linguistics & Phonetics.* vol. 30, no. 3-5, pp. 249-276.

MCALLISTER BYUN, T., HALPIN, P.F. and SZEREDI, D., 2015. Online crowdsourcing for efficient rating of speech: a validation study. *Journal of Communication Disorders.* Jan-Feb, vol. 53, pp. 70-83.

MCALLISTER BYUN, T. and HITCHCOCK, E.R., 2012. Investigating the Use of Traditional and Spectral Biofeedback Approaches to Intervention for /r/ Misarticulation. *American Journal of Speech-Language Pathology.* vol. 21, no. 3, pp. 207-221.

MCALLISTER BYUN, T.M., HITCHCOCK, E.R. and SWARTZ, M.T., 2014. Retroflex versus bunched in treatment for rhotic misarticulation: evidence from ultrasound biofeedback intervention. *Journal of Speech, Language, and Hearing Research : JSLHR.* Dec, vol. 57, no. 6, pp. 2116.

MCCABE, R. and BRADLEY, D., 1975. Systematic multiple phoneme approach to articulation therapy. *Acta Symbolica*. vol. 6, pp. 1–18.

MCGURK, H. and MACDONALD, J., 1976. Hearing lips and seeing voices. *Nature.* Dec 23-30, vol. 264, no. 5588, pp. 746-748.

MCLEOD, S. and BAKER, E., 2017. *Children's speech: an evidence-based approach to assessment and intervention.* Boston: Pearson.

MCLEOD, S., HARRISON, L.J. and MCCORMACK, J., 2012. The intelligibility in Context Scale: validity and reliability of a subjective rating measure. *Journal of Speech, Language, and Hearing Research: JSLHR.* Apr, vol. 55, no. 2, pp. 648-656.

MCREYNOLDS, L.V. and THOMPSON, C.K., 1986. Flexibility of single-subject experimental designs. Part I: Review of the basics of single-subject designs. *The Journal of Speech and Hearing Disorders.* Aug, vol. 51, no. 3, pp. 194-203.

MCWILLIAMS, B.J., 1991. Submucous clefts of the palate: how likely are they to be symptomatic? *The Cleft Palate-Craniofacial Journal.* Jul, vol. 28, no. 3, pp. 251.

MEINUSCH, M. and NEUMANN, S., 2016. Speech and language therapy interventions for children with cleft palate: Evidence not proven. *Evidence-Based Communication Assessment and Intervention.* Oct, vol. 10, no. 3-4, pp. 155.

MELO, R.M., DIAS, R.F., MOTA, H.B. and MEZZOMO, C.L., 2016. Ultrasound images of the tongue prior and post speech therapy. *Revista CEFAC*. Jan/Feb, vol. 18, no. 1, pp.286-297.

MENG, L., BIAN, Z., TORENSMA, R. and VON DEN HOFF, J.W., 2009. Biological mechanisms in palatogenesis and cleft palate. *Journal of Dental Research.* Jan, vol. 88, no. 1, pp. 22-33.

MICHI, K., YAMASHITA, Y., IMAI, S. and OHNO, K., 1990. Results of treatment of speech disorders in cleft palate patients: patients obtaining adequate velopharyngeal function. In G. PFEIFER, ed. *Craniofacial Abnormalities and Clefts of the Lip, Alveolus and Palate.* Stuttgart: Thieme, pp. 419–423.

MICHI, K., YAMASHITA, Y., IMAI, S., SUZUKI, N. and YOSHIDA, H., 1993. Role of visual feedback treatment for defective /s/ sounds in patients with cleft palate. *Journal of Speech and Hearing Research.* Apr, vol. 36, no. 2, pp. 277-285.

MICROTRONICS, CORP., 2016. *PERCI-SAR System, Version 4.01*. [online] [viewed 01 October 2017]. Available from: http://www.microtronics-nc.com/Perci-sar.htm.

MIGUEL, H.C., GENARO, K.F. and TRINDADE, I.E.K., 2007. Perceptual and instrumental assessment of velopharyngeal function in asymptomatic submucous cleft palate. *Pro-Fono: Revista De Atualizacao Cientifica.* Jan-Apr, vol. 19, no. 1, pp. 105-112.

MODHA, G., BERNHARDT, B., CHURCH, R. and BACSFALVI, P., 2008. Case study using ultrasound to treat /r/. *International Journal of Language & Communication Disorders.* vol. 43, no. 3, pp. 323-329.

MORIARTY, B.C. and GILLON, G.T., 2006. Phonological awareness intervention for children with childhood apraxia of speech. *International Journal of Language & Communication Disorders.* Nov-Dec, vol. 41, no. 6, pp. 713-734.

MORLEY, M.E., 1970, *Cleft palate and speech, 7th ed*. Edinburgh: Churchill Livingstone.

MORRIS, H.L., 1973. Velopharyngeal competence and primary cleft palate surgery, 1960-1971: a critical review. *The Cleft Palate Journal.* Jan, vol. 10, pp. 62-71.

MORRIS, H. and OZANNE, A., 2003. Phonetic, phonological, and language skills of children with a cleft palate. *The Cleft Palate-Craniofacial Journal.* Sep, vol. 40, no. 5, pp. 460-470.

MOSSEY, P.A, CASTILLA E., eds., 2003. *Global Registry and Database on Craniofacial Anomalies: Report of a WHO Registry Meeting on Craniofacial Anomalies*. Geneva, Bauru, Brazil: World health Organization.

MOSSEY, P.A., LITTLE, J., MUNGER, R.G., DIXON, M.J. and SHAW, W.C., 2009. Cleft lip and palate. *Lancet (London, England).* Nov 21, vol. 374, no. 9703, pp. 1773-1785.

MUNSON, B., SCHELLINGER, S.K. and CARLSON, K.U., 2012. Measuring Speech-Sound Learning Using Visual Analog Scaling. *SIG 1 Perspectives on Language Learning and Education.* Jan, vol. 19, no. 1, pp. 19-30.

MURRAY, E., MCCABE, P. and BALLARD, K.J., 2015. A Randomized Controlled Trial for Children With Childhood Apraxia of Speech Comparing Rapid Syllable Transition Treatment and the Nuffield Dyspraxia Programme-Third Edition. *Journal of Speech, Language, and Hearing Research: JSLHR.* Jun, vol. 58, no. 3, pp. 669-686.

NATIONAL LITERACY TRUST (NLT)., 2014. *Parents' Perspectives: Children's Use of Technology in the Early Years.* [online] [viewed 05 August 2017]. Available from:

http://www.literacytrust.org.uk/assets/0002/1140/Early_years_parent_report.pdfhttp://www.literacytrust.org.uk/ assets/0002/1140/Early_years_parent_report.pdf.

NEUMANN, S. and ROMONATH, R., 2012. Effectiveness of nasopharyngoscopic biofeedback in clients with cleft palate speech: a systematic review. *Logopedics, Phoniatrics, Vocology.* Oct, vol. 37, no. 3, pp. 95-106.

O'NEILL, G., 2004. *The use of a new computer programme in the treatment of dysarthria: A single case study.* Bachelor Thesis, Trinity College Dublin.

OFCOM., 2014. *Children and Parents: Media Use and Attitudes Report*. [online] [viewed 05 August 2017]. Available from: https://www.ofcom.org.uk/__data/assets/pdf_file/0027/76266/childrens_2014_report.pdf.

PAMPLONA, C., YSUNZA, A., PATIÑO, C., RAMÍREZ, E., DRUCKER, M. and MAZÓN, J.J., 2005. Speech summer camp for treating articulation disorders in cleft palate patients. *International Journal of Pediatric Otorhinolaryngology.* Mar, vol. 69, no. 3, pp. 351-359.

PAMPLONA, M.C., YSUNZA, A. and ESPINOSA, J., 1999. A comparative trial of two modalities of speech intervention for compensatory articulation in cleft palate children, phonologic approach versus articulatory approach. *International Journal of Pediatric Otorhinolaryngology.* Jun, vol. 49, no. 1, pp. 21-26.

PASSY, J., 2010. *Cued articulation: consonants and vowels.* Camberwell, Vic.: ACER Press.

PATERSON, P., SHER, H., WYLIE, F., WALLACE, S., CRAWFORD, A., SOOD, V., GILLGRASS, T., RAY, A. and DEVLIN, M., 2011. Cleft lip/palate: incidence of prenatal diagnosis in Glasgow, Scotland, and comparison with other centers in the United Kingdom. *The Cleft Palate-Craniofacial Journal.* Sep, vol. 48, no. 5, pp. 608-613.

PEELLE, J.E. and SOMMERS, M.S., 2015. Prediction and constraint in audiovisual speech perception. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior.* Jul, vol. 68, pp. 169-181.

PERSSON, C. and SJOGREEN, L. 2011. The Influence of Related Conditions on Speech and Communication. In: S. HOWARD and A. LOHMANDER, eds. *Cleft palate speech: assessment and intervention.* Chichester: Wiley-Blackwell, pp. 41-53.

PETERSON-FALZONE, S.J., TROST-CARDAMONE, J.E., KARNELL, M.P. and HARDIN-JONES, M.A., 2006. *The Clinician's guide to treating cleft palate speech.* St. Louis, Miss: Mosby Elsevier.

PETERSON-FALZONE, S.J., HARDIN-JONES, M.A. and KARNELL, M.P., 2010. *Cleft palate speech.* St. Louis, Mo: Mosby/Elsevier.

PRATHANEE, B., PUMNUM, T. and SEEPUAHAM, C., 2012. Relationship of perceptual evaluation for resonance disorders to nasalance scores in children with cleft palate. *Journal of the Medical Association of Thailand = Chotmaihet Thangphaet.* Nov, vol. 95, pp. 100.

PRESTON, J.L., BRICK, N. and LANDI, N., 2013. Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology.* Nov, vol. 22, no. 4, pp. 627.

PRESTON, J.L. and EDWARDS, M.L., 2007. Phonological processing skills of adolescents with residual speech sound errors. *Language, Speech, and Hearing Services in Schools.* Oct, vol. 38, no. 4, pp. 297-308.

PRESTON, J.L. and LEAMAN, M., 2014. Ultrasound visual feedback for acquired apraxia of speech: A case report. *Aphasiology.* Mar, vol. 28, no. 3, pp. 278-295.

PRESTON, J.L., LEECE, M.C. and MAAS, E., 2016a. Intensive Treatment with Ultrasound Visual Feedback for Speech Sound Errors in Childhood Apraxia. *Frontiers in Human Neuroscience.* vol. 10, pp. 440.

PRESTON, J.L., LEECE, M.C. and MAAS, E., 2017a. Motor-based treatment with and without ultrasound feedback for residual speech-sound errors. *International Journal of Language & Communication Disorders.* Jan, vol. 52, no. 1, pp. 80-94.

PRESTON, J.L., LEECE, M.C., MCNAMARA, K. and MAAS, E., 2017b. Variable Practice to Enhance Speech Learning in Ultrasound Biofeedback Treatment for Childhood Apraxia of Speech: A Single Case Experimental Study. *American Journal of Speech-Language Pathology.* Aug, vol. 26, pp. 840-852.

PRESTON, J.L., MAAS, E., WHITTLE, J., LEECE, M.C. and MCCABE, P., 2016b. Limited acquisition and generalisation of rhotics with ultrasound visual feedback in childhood apraxia. *Clinical Linguistics & Phonetics.* vol. 30, no. 3-5, pp. 363-381.

PRESTON, J., MCCABE, P., RIVERA-CAMPOS, A., WHITTLE, J., LANDRY, E. and MAAS, E., 2014. Ultrasound Visual Feedback Treatment and Practice Variability for Residual Speech Sound Errors. *Journal of Speech Language and Hearing Research.* DEC, vol. 57, no. 6, pp. 2102-2115.

PRESTON, J.L., RAMSDELL, H.L., OLLER, D.K., EDWARDS, M.L. and TOBIN, S.J., 2011. Developing a weighted measure of speech sound accuracy. *Journal of Speech, Language, and Hearing Research: JSLHR.* Feb, vol. 54, no. 1, pp. 1-18.

PRO-ED, LTD., 1986. *See-Scape. Visual Feedback of Nasal Emission*. Austin, Texas: Pro-Ed, Ltd.

RAHIMOV, F., JUGESSUR, A. and MURRAY, J.C., 2012. Genetics of nonsyndromic orofacial clefts. *The Cleft Palate-Craniofacial Journal.* Jan, vol. 49, no. 1, pp. 73-91.

RAVEN, J., RAVEN, J. and COURT, J., 1998. *Raven's Progressive Matrices and Raven's Coloured Matrices.* Oxford, England: Oxford Psychologists Press.

REID, E., 1978. Social and stylistic variation in the speech of children: some evidence from Edinburgh. In P. TRUDGILL, ed. *Sociolinguistic Patterns in British English*. London: Arnold, pp. 158-171.

REITER, R., BROSCH, S., WEFEL, H., SCHLÖMER, G. and HAASE, S., 2011. The submucous cleft palate: diagnosis and therapy. *International Journal of Pediatric Otorhinolaryngology.* Jan, vol. 75, no. 1, pp. 85-88.

RICHSTSMEIER, P., 2010. *Child phoneme errors are not substitutions.* Toronto Working Papers in Linguistics (TWPL). vol. 33.

RISKI, J.E. and DELONG, E., 1984. Articulation development in children with cleft lip/palate. *The Cleft Palate Journal.* Apr, vol. 21, no. 2, pp. 57-64.

ROXBURGH, Z., CLELAND, J. and SCOBBIE, J.M., 2016. Multiple phonetically trained-listener comparisons of speech before and after articulatory intervention in two children with repaired submucous cleft palate. *Clinical Linguistics & Phonetics.* vol. 30, no. 3-5, pp. 398-415.

ROYAL COLLEGE OF SPEECH AND LANGUAGE THERAPISTS (RCSLT)., 2005. *RCSLT Clinicial Guidelines.* Oxon: Speechmark Publishing.

RUSCELLO, D. and VALLINO, L., 2014. The Application of Motor Learning Concepts to the Treatment of Children with Compensatory Speech Sound Errors. *SIG 5 Perspectives on Speech Science and Orofacial Disorders.* Oct, vol. 24, pp. 39-47.

RUSSELL, V.J., GRUNWELL, P., 1993. Speech development in children with cleft lip and palate. In: P. Grunwell, ed. *Analysing Cleft Palate Speech*. London: Whurr Publishers, pp. 19–47.

RVACHEW, S. and BROSSEAU-LAPRE, F., 2012. *Developmental Phonological Disorders: Foundations of Clinical Practice.* San Diego, CA: Plural Publishing, Co.

SANDY, J., WILLIAMS, A., MILDINHALL, S., MURPHY, T., BEARN, D., SHAW, B., SELL, D., DEVLIN, B. and MURRAY, J., 1998. The Clinical Standards Advisory Group (CSAG) Cleft Lip and Palate Study. *British Journal of Orthodontics.* Feb, vol. 25, no. 1, pp. 21-30.

SATAKE, E., JAGAROO, V. and MAXWELL, D.L., 2008. *Handbook of statistical methods: single subject design.* San Diego, Oxford, Brisbane: Plural Publishing Inc.

SCHELLINGER, S.K., MUNSON, B. and EDWARDS, J., 2017. Gradient perception of children's productions of /s/ and /θ/: A comparative study of rating methods. *Clinical Linguistics & Phonetics.* Jan 2, vol. 31, no. 1, pp. 80-103.

SCHERER, N. and LOUW, B., 2011. Early Communication Assessment and Intervention. In: S. HOWARD and A. LOHMANDER, eds. *Cleft palate speech: assessment and intervention.* Chichester: Wiley-Blackwell, pp. 259-274.

SCHMIDT, R. and LEE, T., 2005. *Motor Control and Learning: A behavioural emphasis. 4th ed*. Champaign, IL: Human Kinetics.

SCOBBIE, J.M., LAWSON, E., COWEN, S., CLELAND, J. and WRENCH, A.A., 2011. *A common co-ordinate system for mid-sagittal articulatory measurement.* QMU CASL Research Centre: Working Paper WP-20.

SCOBBIE, J.M., LAWSON, E. and STUART-SMITH, J., 2012. Back to front: a socially-stratified ultrasound tongue imaging study of Scottish English /u/. *Rivista Di Linguistica / Italian Journal of Linguistics, Special Issue: "Articulatory Techniques for Sociophonetic Research".* vol. 24, no. 1, pp. 103-148.

SCOBBIE, J. and J. CLELAND., 2017. 7th International Conference on Speech Motor Control. *Quantitative measurement of the dorsal constriction in the /k/-/t/ contrast and its co-articulatory variation in children and adults* [poster]. Groningen, Netherlands.

SELL, D., 2005. Issues in perceptual speech analysis in cleft palate and related disorders: a review. *International Journal of Language & Communication Disorders.* Apr, vol. 40, no. 2, pp. 103-121.

SELL D. and GRUNWELL, P.A., 2001. Speech assessment and therapy. In: A.C.H., WATSON, D. SELL and P. GRUNWELL, eds. *Management of Cleft Lip and Palate*. London: Whurr, pp.258-285.

SELL, D., HARDING, A. and GRUNWELL, P., 1994, GOS.SP.ASS. A screening assessment of cleft palate speech. *European Journal of Disorders of Communication*. vol. 29, pp. 1–15.

SELL, D., HARDING, A. and GRUNWELL, P., 1999. GOS.SP.ASS.'98: an assessment for speech disorders associated with cleft palate and/or velopharyngeal dysfunction (revised). *International Journal of Language & Communication Disorders.* Jan, vol. 34, no. 1, pp. 17-33.

SELL, D., JOHN, A., HARDING-BELL, A., SWEENEY, T., HEGARTY, F. and FREEMAN, J., 2009. Cleft audit protocol for speech (CAPS-A): a comprehensive training package for speech analysis. *International Journal of Language & Communication Disorders.* Jul-Aug, vol. 44, no. 4, pp. 529-548.

SELL, D., MILDINHALL, S., MURPHY, T., CORNISH, T. and GRUNWELL, P., 2002. Speech analysis: audio or video recordings? [Conference Presentation]. *Craniofacial Society of Great Britain and Ireland*. East Grinstead, UK.

SEMEL, E., WIIG, E.H. and SECORD, W.A., 2006. *Clinical Evaluation of Language Fundamentals 4th ed. (CELF-4).* London: Pearson Assessment.

SHARP, H.M., DAILEY, S. and MOON, J.B., 2003. Speech and language development disorders in infants and children with cleft lip and palate. *Pediatric Annals.* Jul, vol. 32, no. 7, pp. 476-480.

SHAWKER, T.H. and SONIES, B.C., 1985. Ultrasound biofeedback for speech training. Instrumentation and preliminary results. *Investigative Radiology.* Jan-Feb, vol. 20, no. 1, pp. 90-93.

SHPRINTZEN, R.J., SCHWARTZ, R.H., DANILLER, A. and HOCH, L., 1985. Morphologic significance of bifid uvula. *Pediatrics.* Mar, vol. 75, no. 3, pp. 553-561.

SHPRINTZEN, R.J., 2008. Velo-cardio-facial syndrome: 30 Years of study. *Developmental Disabilities Research Reviews.* Jan, vol. 14, no. 1, pp. 3-10.

SHRIBERG, L.D., 1980. An intervention procedure for children with persistent /r/ errors. *Language, Speech, and Hearing Services in Schools*. vol. 11, no. 2, pp. 102-110.

SHRIBERG, L.D., 1982. Diagnostic assessment of developmental phonological disorders. In M. CRARY, ed. *Phonological intervention, concepts and procedures*. San Diego: College-Hill Inc.

SHRIBERG, L.D., AUSTIN, D., LEWIS, B.A., MCSWEENY, J.L. and WILSON, D.L., 1997. The speech disorders classification system (SDCS): extensions and lifespan reference data. *Journal of Speech, Language, and Hearing Research: JSLHR.* Aug, vol. 40, no. 4, pp. 723-740.

SHRIBERG, L.D. and KWIATKOWSKI, J., 1980. *Natural Process Analysis: A procedure for phonological analysis of continuous speech samples*. New York: Macmillan.

SHRIBERG, L.D. and KWIATKOWSKI, J., 1982. Phonological disorders III: a procedure for assessing severity of involvement. *The Journal of Speech and Hearing Disorders.* Aug, vol. 47, no. 3, pp. 256-270.

SHRIBERG, L.D., KWIATKOWSKI, J., BEST, S., TERSELIC-WEBER, B. and HENGST, J., 1986. Characteristics of Children with Phonologic Disorders of Unknown Origin. *Journal of Speech and Hearing Disorders.* vol. 51, no. 2, pp. 140-161.

SHRIBERG, L.D. and LOF, G.L., 1991. Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics & Phonetics.* Jan, vol. 5, no. 3, pp. 225-279.

SJOLIE, G., 2015. *Effects of Ultrasound as Visual Feedback of the Tongue on Generalization, Retention, and Acquisition in Speech Therapy for Rhotics.* Master of Science Thesis. Syracuse University.

SJOLIE, G.M., LEECE, M.C. and PRESTON, J.L., 2016. Acquisition, retention, and generalization of rhotics with and without ultrasound visual feedback. *Journal of Communication Disorders.* Nov-Dec, vol. 64, pp. 62-77.

SKELTON, S.L., 2004. Motor skill learning approaches to the treatment of speech sound disorders. *California Speech Hearing Association Magazine*, *Summer*, pp.8-9.

SKELTON, S.L. and HAGOPIAN, A.L., 2014. Using randomized variable practice in the treatment of childhood apraxia of speech. *American Journal of Speech-Language Pathology.* Nov, vol. 23, no. 4, pp. 599-611.

SMARTY EARS., 2011. *Speech Trainer 3D.* [online] [viewed 01 November 2011]. Available from: http://smartyearsapps.com.

SMITH, J, and HOLMES-ELLIOTT, S., 2017, 'The unstoppable glottal: tracking rapid change in an iconic British variable[1], *English Language and Linguistics*, pp. 1-33.

SPRIESTERSBACH, D.C., DARLEY, F.L. and ROUSE, V., 1956. Articulation Of A Group Of Children With Cleft Lips And Palates. *Journal of Speech and Hearing Disorders.* vol. 21, no. 4, pp. 436-445.

STACKHOUSE, J. and WELLS, B., 1997. *Children's speech and literacy difficulties.* London: Whurr.

STENGELHOFEN, J., 1989. *Cleft palate: the nature and remediation of communication problems.* Edinburgh; New York: Churchill Livingstone.

STEWART, J.M., OTT, J.E. and LAGACE, R., 1972. Submucous cleft palate: prevalence in a school population. *The Cleft Palate Journal.* Jul, vol. 9, pp. 246-250.

STOLL, C., ALEMBIK, Y., DOTT, B. and ROTH, M.P., 2000. Associated malformations in cases with oral clefts. *The Cleft Palate-Craniofacial Journal.* Jan, vol. 37, no. 1, pp. 41-47.

STONE, M., 1999. Laboratory Techniques for Investigating Speech Articulations. In: W.J., HARDCASTLE and J. LAVER, eds. *The Handbook of Phonetic Sciences*. Oxford; Blackwell Publishers, pp.11-32.

STONE, M., FABER, A., RAPHAEL, L. J. and SHAWKER, T. H., 1992, Cross-sectional tongue shape and linguapalatal contact patterns in [s], [b], and [l]. *Journal of Phonetics*. vol. 20, pp. 253–270.

SULLIVAN, S.R., VASUDAVAN, S., MARRINAN, E.M. and MULLIKEN, J.B., 2011. Submucous cleft palate and velopharyngeal insufficiency: comparison of speech outcomes using three operative techniques by one surgeon. *The Cleft Palate-Craniofacial Journal.* Sep, vol. 48, no. 5, pp. 561-570.

THOMPSON, E., 2015. *Intelligibility pre/post therapy in children with cleft lip and palate*. Unpublished Honours Project, Queen Margaret University, Edinburgh.

TREILLE, A., VILAIN, C., HUEBER, T., SCHWARTZ, J.L., LAMALLE, L. and SATO, M., 2014. Inside speech: neural correlates of audio-lingual speech perception. *Neurobiology of Language Conference*, 27-29 August, Amsterdam, The Netherlands.

TROST, J.E., 1981. Articulatory additions to the classical description of the speech of persons with cleft palate. *The Cleft Palate Journal.* Jul, vol. 18, no. 3, pp. 193-203.

ULTRAPHONIX., 2015-2016. UltraPhonix: ultrasound visual Biofeedback Treatment for Speech Sound Disorders in Children. *CSO (Chief Scientist Office)*. Grant number ETM/402.

ULTRAX., 2011-2014. ULTAX: Real-time tongue tracking for speech therapy using ultrasound. *EPSRC (Engineering and Physical Sciences Research Council)*. Healthcare Partnership research grant EP/I027696/1.

ULTRAX2020., 2017-2020. Ultrax2020: Ultrasound Technology for Optimising the Treatment of Speech Disorders. *EPSRC (Engineering and Physical Sciences Research Council)*. Healthcare Partnership research grant EP/I027696/1.

VALLINO-NAPOLI, L.D., 2011. Evaluation and Evidence-Based Practice. In: S. HOWARD and A. LOHMANDER, eds. *Cleft palate speech: assessment and intervention.* Chichester: Wiley-Blackwell, pp. 317-358.

VAN BORSEL, J., REUNES, G. and VAN DEN BERGH, N., 2003. Delayed auditory feedback in the treatment of stuttering: clients as consumers. *International Journal of Language & Communication Disorders.* Apr, vol. 38, no. 2, pp. 119-129.

VAN DENMARK, D.R., 2004. Speech and Voice Therapy Techniques for School-Age and Adult Patients with Remaining Cleft Palate Speech Disorders. In: K.R., BZOCH, ed. *Communicative Disorders Related To Cleft Lip and Palate*. Austin, Texas: Pro-ed, pp.741-762.

VAN DEMARK, D.R. and HARDIN, M.A., 1986. Effectiveness of intensive articulation therapy for children with cleft palate. *The Cleft Palate Journal.* Jul, vol. 23, no. 3, pp. 215-224.

VAN LIERDE, K.M., DE BODT, M., VAN BORSEL, J., WUYTS, F.L. and VAN CAUWENBERGE, P., 2002. Effect of cleft type on overall speech intelligibility and resonance. *Folia Phoniatrica Et Logopaedica.* May-Jun, vol. 54, no. 3, pp. 158-168.

VAN RIPER, C., 1978. *Speech correction: principles and methods.* Englewood Cliffs, N.J.: Prentice-Hall.

VAN RIPER, C. and EMERICK, L., 1984. *Speech correction: an introduction to speech pathology and audiology. 7th ed*. Englewood Cliffs, N.J: Prentice-Hall.

WATSON, A.C.H., GRUNWELL, P. and SELL, D.A., 2000. *Management of cleft lip and palate.* London: Whurr.

WEATHERLEY-WHITE, R.C., SAKURA, C.Y., BRENNER, L.D., STEWART, J.M. and OTT, J.E., 1972. Submucous cleft palate. Its incidence, natural history, and indications for treatment. *Plastic and Reconstructive Surgery.* Mar, vol. 49, no. 3, pp. 297-304.

WEVOSYS., 2017. *LingWAVES, A leading tested product for voice and speech analysis, biofeedback and documentation*. [online] [viewed 01 October 2017]. Available from: https://www.wevosys.com/products/lingwaves/lingwaves.html.

WHARTON, P. and MOWRER, D.E., 1992. Prevalence of cleft uvula among school children in kindergarten through grade five. *The Cleft Palate-Craniofacial Journal.* Jan, vol. 29, no. 1, pp. 14.

WIGHTMAN, D. C. and LINTERN, G., 1985. Part-task training for tracking and manual control. *Human Factors,* vol. 27 no. *3*, pp. 267-283.

WOO, A.S., 2012. Velopharyngeal dysfunction. *Seminars in Plastic Surgery.* Nov, vol. 26, no. 4, pp. 170-177.

WOOD, S., CLELAND, J. and ROXBURGH, Z., 2015. Powerful tools for motor-based treatment approaches. *Royal College of Speech and Language Therapists Bulletin.* vol. 762, pp. 18-20.

WRENCH. A. and SCOBBIE, J.M., 2016. *Queen Margaret University ultrasound, audio and video multichannel recording facility (2008-2016).* QMU CASL Research Centre: Working Paper WP-24.

WYATT, R., SELL, D., RUSSELL, J., HARDING, A., HARLAND, K. and ALBERY, L., 1996. Cleft palate speech dissected: a review of current knowledge and analysis. *British Journal of Plastic Surgery.* January 1, vol. 49, no. 3, pp. 143-149.

YAMASHITA, Y. and MICHI, K., 1991. Misarticulation caused by abnormal lingual-palatal contact in patients with cleft palate with adequate velopharyngeal function. *The Cleft Palate-Craniofacial Journal.* Oct, vol. 28, no. 4, pp. 368.

YIU, E.M. and NG, C., 2004. Equal appearing interval and visual analogue scaling of perceptual roughness and breathiness. *Clinical Linguistics & Phonetics.* Apr, vol. 18, no. 3, pp. 211-229.

YOUNG, E., 2015. *The Effectiveness of Ultrasound Visual Biofeedback in the remediation of an Idiosyncratic Case of Alveolar Backing*. Unpublished Master's Thesis, Queen Margaret University, Edinburgh.

YOUSSEF, A.B., HUEBER, T., BADIN, P. and BAILLY, G., 2011 Toward a Multi-Speaker Visual Articulatory Feedback System. *Interspeech 2011* (12th Annual Conference of the International Speech Communication Association), pp. 589-592. Florence, Italy, 28-31 August.

ZHARKOVA, N., 2013. Using Ultrasound to Quantify Tongue Shape and Movement Characteristics. *Cleft Palate-Craniofacial Journal.* Jan, vol. 50, no. 1, pp. 76.

ZHARKOVA, N., HEWLETT, N. and LICKLEY, R., 2012. *END OF AWARD REPORT- ESRC RES.* Swindon: ESRC.

7 Appendices

7.1 Appendix 1: Information Sheets for Treatment Study

Queen Margaret University

EDINBURGH

Looking at speech: can our special computers help your talking?

Information Sheet for Children aged 6-12: To be read to children by parent/carer

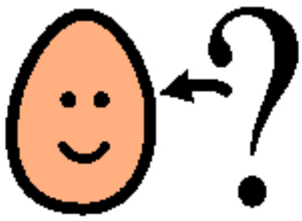
Would you like to take part in an experiment about speaking?

• Please read the information below carefully and ask about anything you don't understand.

• This will help you to make up your mind about whether you would like to take part in the experiment or not.

• If you have any questions or worries, please ask the adult who looks after you, or feel free to contact **Zoe Roxburgh on 0131 474 0000** (Say "Zoe Roxburgh" when prompted by the automated voice system).

Ask me questions



• This experiment will look at pictures of a mouth on an iPad and we will look at the patterns your tongue makes when you talk by showing this on a computer screen so we can help you to speak more clearly and easily.

WHAT DO I HAVE TO DO IF I TAKE PART?

• You will have a special headset/helmet to wear so we can record how your tongue moves when you say different things.

A picture of the headset

• We will use a special microphone and computer to record you speaking out loud when you are wearing the headset.

A picture of the special computer

• You will use an iPad to show you pictures of a tongue moving.

A picture of the iPad

• You will be asked to say some words and to have a chat with the speech and language therapist. We will ask you to look at some pictures and talk about them with the speech and language therapist.

• You will be asked to answer some questions about speaking.

• We will ask you to come with your Mum, Dad or person who looks after you to our clinic at Queen Margaret University to make these recordings because the special computer can't be moved.

• We will ask you to come for 6 recordings on 6 different days.

• We will use the special computer and the iPad to help you with your speaking. You will come to Queen Margaret University once a week for 8weeks for speech therapy, using the computer and once a week for 8weeks for speech therapy using the iPad.

Queen Margaret University

WHAT HAPPENS AFTER I TAKE PART?

• We will keep the recordings we made on the computer and the recording we made with the video camera.

- Other people will listen to your voice to decide if your speaking has got better after therapy at Queen Margaret University.

• We will write to your mum, dad or person who looks after you to tell you what we found out in our research project. You can come back to visit us to see what we found out.

Queen Margaret University
EDINBURGH

Visualising speech: can ultrasound or Speech Trainer 3D treat speech disorders associated with repaired cleft palate?

Information Sheet for Parents/ Carers of Potential Participants

Your child is being invited to take part in a research study. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve.

Please take time to read the following information carefully. Talk to others about the study if you wish. Contact us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish for your child to take part.

What is the purpose of the study?
The aim of the research is to test and compare two tools used in speech therapy for treating speech sound disorders associated with repaired cleft palate. This study will use standard medical ultrasound to record the movements of the tongue during speech and the app Speech Trainer 3D on an iPad3 to demonstrate speech movements. We want to find out if ultrasound and Speech Trainer 3D can be used to help children with speech disorders, as a result of repaired cleft palate, improve their speech. The ultrasound technology will allow your child to view the movements of his or her own tongue whilst s/he is speaking and to modify these movements in order to help with any difficulties s/he may be having producing speech sounds. Speech Trainer 3D provides animations of speech movements.

Why has my child been asked to take part?
Your child has been invited to take part because he or she has been identified by his or her Speech and Language Therapist as potentially suitable for our research project because he/she has previously been diagnosed with a Speech Sound Disorder as a result of a repaired cleft palate, is aged between 6 and 16 and lives within Edinburgh, Lothians or Fife.

English is required as a first language for this study. This is because it has a focus on treating English speech sounds. The app Speech Trainer 3D also only uses only English speech sounds.

Does my child have to take part?

No, it is up to you to decide whether or not your child takes part. If you do decide for your child to take part you will be given this information sheet to keep and will be asked to sign a consent form. If you decide for your child to take part you are still free to withdraw at any time and without giving a reason. However, if you decide to withdraw your child from the study, anonymised data that has already been stored and analysed will not be destroyed. Deciding not to take part or withdrawing from the study will not affect the healthcare or speech and language therapy your child receives.

What will happen if my child takes part?

• Your child will be asked to come to Queen Margaret for a speech assessment for us to check whether or not they are suitable. If so, we will offer two blocks of 8 x 1 hour sessions of therapy as part of the study and up to 6 further assessment sessions.

• During some of the sessions your child will be asked to sit in front of a computer screen in a sound-treated studio. Your child will use a headset, which will ensure that the ultrasound probe can be correctly positioned beneath the chin. The end of the probe will be covered in medical gel. A microphone and video camera will be attached to the helmet, in order to record the voice of your child and a video of your child's lips when s/he speaks. During the other sessions your child will be asked to sit at a table in a clinic room to use the iPad3. The Speech and Language Therapist will be with your child during each session and will answer any questions you have throughout.

• Your child will be asked to copy various sounds, words and sentences and possibly drink a few sips of water. He or she will also be asked to look at the live image of their own tongue and reflect on this. The sessions will be recorded for analysis. Your child will also be asked to try to copy images on the iPad3. You and your child will also be asked to complete questionnaires.

• Each session will take about 1 hour, including rest breaks and fun games.

• It will be necessary for both you and your child to travel to Queen Margaret University for the assessment and therapy session included in this study.

• In addition, in order to obtain relevant background information, your child will be given speech and language assessments.

• With your permission we will consult with your child's Speech and Language Therapist and report results of the study back to her. With your permission we will inform you child's GP that he or she is taking part in the research project.

All data will be anonymised. Your child will not be mentioned by name in any report or presentation. However, if some of the recordings or videos were played at a verbal presentation or for teaching purposes, there is the possibility that the voice of your child may be recognisable or that they may be identified through a video recording, due to there being a wider audience.

Recordings of your child's voice will be played to a number of people as part of a listening experiment. The purpose of this listening experiment is to explore how unfamiliar listeners perceive speech and to achieve a more accurate judgement on whether their speech has improved after speech therapy.

This study is an addition to any Speech and Language Therapy your child is receiving on the NHS. Therefore, NHS treatment will not be affected by participating.

What are the possible benefits of taking part?

If your child takes part in the project they will have the benefit of an in-depth speech and language assessment and two courses of speech therapy which may or may not help them with his or her speech disorder. The possibility of your child's speech deteriorating is negligible.

What are the possible disadvantages and risks of taking part?

It is not thought that there are many disadvantages and Ultrasound is subject to rigorous safety assessments. At all levels of intensity used for diagnostic imaging, there are no known risks associated with ultrasound and there are no specific dangers or safety requirements. The ultrasound equipment and headset has been used before at Queen Margaret University with both children and adults. Your child may experience some mild discomfort from wearing the headset. We will provide rest breaks as required and the experiment can be discontinued at any point if you or your child wishes. There are also no known risks associated with the Speech Trainer 3D app for the iPad3.

One potential disadvantage of taking part is the inconvenience of travelling to Queen Margaret
University for up to 22 separate visits as unfortunately it will not be possible for the ultrasound equipment to be moved to your home or your child's school. A £20

Amazon gift voucher will be offered for taking part and a contribution to travel expenses will be made.

What happens when the study is finished?

We will write to you within two months with a report detailing your child's individual speech skills and their individual progress with the ultrasound therapy and therapy using Speech Trainer 3D carried out during the project. With your permission we will share this information with your child's speech and language therapist. Due to the time limited nature of PhD funding, it is unlikely that we will be able to offer your child further therapy and you should be aware that ultrasound therapy is not available from the NHS due to its experimental nature.

Will taking part in the study be kept confidential?

All the information we collect during the course of the research will be kept confidential and there are strict laws which safeguard your privacy at every stage. Your child's name will be removed from the data so that s/he cannot be recognised from it. Data will be kept at Queen Margaret University. With your consent we will inform your child's Speech and Language Therapist that you are taking part.

What will happen to the results of the study?

The results of the study will be shared with the public, Speech and Language Therapists and academics via our website. It will also be used for teaching purposes, conference presentations and publication in academic journals. We will invite you and your child to attend an information day to tell you about the results of the study towards the end of the project (Late 2014).

Who is organising the research and why?

This study has been organised by Queen Margaret University as part of a PhD studentship and is funded by Queen Margaret University. It will run from September 2011 to September 2014.

Who has reviewed the study?

The study proposal has been reviewed by independent reviewers at Queen Margaret University as part of a probationary assessment. A favourable ethical opinion has been obtained from South East Scotland REC 01 and Queen Margaret University ethics committee. NHS management approval has also been obtained.

It is entirely up to you to decide whether or not your child takes part in the project. If you do decide for your child to participate, you will be given this information sheet to keep and be asked to sign a consent form. You and your child are free to withdraw from the study at any stage without giving a reason.

If you would like to consult an independent person, who knows about this project but is not involved in it, you are welcome to contact Dr Janet Beck, 0131 474 0000 (Say “Janet Beck” when prompted by the automated voice system). You are also welcome to contact my PhD supervisor, Dr Joanne Cleland, an experienced Speech and Language Therapist, on 0131 474 0000 (Say “Joanne Cleland” when prompted by the automated voice system) or email JCleland@qmu.ac.uk.

If you have read and understood this information sheet, and you think you might be interested in having your child participating in the study, please now email ZRoxburgh@qmu.ac.uk. There will be an opportunity to ask questions and sign the consent form when you come to Queen Margaret University to meet with the research SLT/ PhD Student, Zoe Roxburgh.

Thank you for taking the time to read this information.

Contact details of the researcher:
Name of PhD Student/ Speech and Language Therapist: Zoe Roxburgh

Address:
Speech and Hearing Sciences,
Queen Margaret University, Edinburgh
Queen Margaret University Drive
Musselburgh
East Lothian EH21 6UU
Email / Telephone: ZRoxburgh@qmu.ac.uk / 0131 474 0000 (Say “Zoe Roxburgh” when prompted by the automated voice system)

If you wish to make a complaint about the study please contact NHS Lothian:
NHS Lothian Complaints Team
2nd Floor
Waverley Gate
2-4 Waterloo Place
Edinburgh
EH1 3EG
Tel: 0131 465 5708

7.2 Appendix 2: Consent Forms for Therapy Study

Queen Margaret University

EDINBURGH

Participant Consent Form Children Under 12

Project Title: Looking at speech: can our special computers help your talking?

Name of researcher: Zoe Roxburgh

Address: PhD Student, Speech and Hearing Sciences,
School of Health Sciences
Queen Margaret University, Edinburgh
Queen Margaret University Drive
Musselburgh
East Lothian EH21 6UU

Email / Telephone: ZRoxburgh@qmu.ac.uk / 0131 474 0000 (Say "Zoe Roxburgh" when prompted by the automated voice system)

Thank you for reading the information about our research project. If you would like to take part, please read and sign this form.

My name ____________________________________ My age _________

Please Initial Box

1. I have read and understand the information sheet dated 25/09/2013, (Version 4.0) and have had the opportunity to ask questions. ☐

2. I understand that participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. ☐

3. I agree to taking part in this project and that videos, recordings of my voice and photographs of me can be kept and studied by the researchers. My name will not be on anything the researchers keep. ☐

4. I agree that videos and recordings of my voice can be used in

university teaching for people learning how to be Speech Therapists.

5. I agree that videos and recordings of my voice can be played on the internet or at special events that anyone can go to.
6. I agree that videos and recordings of my voice can be copied and given to other researchers outside QMU to study.
7. I agree for the people working on the research project to look at my Speech and Language Therapy notes.
8. I agree for people working on the research project to speak to my usual Speech and Language therapist about my speaking skills.
9. I agree for people working on the research project to write to my Doctor to tell him or her that I am taking part in this project.
10. I agree that recordings of my voice can be used in a listening experiment at QMU where lots of people will listen to my voice.

Name of Participant	Signature	Date
Name of Person taking consent	Signature	Date
Researcher	Signature	Date

Thank you for agreeing to participate in this research.

Queen Margaret University

EDINBURGH

Participant Consent Form Parents/Carers

Project Title: Visualising speech: can ultrasound or Speech Trainer 3D treat speech disorders associated with repaired cleft palate?

Name of researcher: Zoe Roxburgh

Address: PhD Student, Speech and Hearing Sciences,
School of Health Sciences
Queen Margaret University, Edinburgh
Queen Margaret University Drive
Musselburgh
East Lothian EH21 6UU

Email / Telephone: ZRoxburgh@qmu.ac.uk / 0131 474 0000 (Say "Zoe Roxburgh" when prompted by the automated voice system)

Thank you for reading the information about our research project. If you would like to take part, please read and sign this form.

Participant's name ________________________________Date of Birth_________

Please Initial Box

1. I have read and understand the information sheet dated 25/09/2013, (Version 4.0) and have had the opportunity to ask questions. ☐

2. I understand that participation is voluntary and that my child is free to withdraw at any time, without giving any reason, without my child's medical care or legal rights being affected. ☐

3. I agree to my child participating in this study and that any audio and visual ultrasound, video and photographic data can be stored and used indefinitely but anonymously for analysis, research, academic conference presentations, and future applications for research funding, and that the anonymous results of such analyses can be disseminated freely to audiences and research users of all types. ☐

4. I agree for relevant sections of my child's Speech and Language Therapy notes or data collected during the study, to be looked at by individuals from QMU or from the NHS Trust, where it is relevant to taking part in this research. I give permission for these individuals to have access to my child's records.

5. I agree for the research team to contact my child's NHS Speech and Language therapist to discuss my child's speech and language skills and pass on information and results of therapy.

6. I give permission for my child's samples/data to be kept for use in future ethically approved research.

7. I give permission for my child's GP to be informed of my child's participation and given any relevant information.

8. I agree that anonymous recordings of my child's voice and visual images from ultrasound and video can be used in university teaching.

9. I agree that anonymous recordings of my child's voice and visual images from ultrasound and video can be played to a public audience to advance understanding of science, through the internet, broadcast, laboratory open days, science festivals and other public but non-professional talks and presentations.

10. I agree that anonymous recordings of my child's voice can be used in a listening experiment at QMU, where listeners will make a judgement on whether my child's speech has improved after therapy.

Name of Participant	Signature	Date
Name of Parent/Carer	Signature	Date

______________________ Name of Person taking consent	Signature	Date
______________________ Researcher	______________________ Signature	______________________ Date

Thank you for agreeing to participate in this research.

7.3 Appendix 3: Therapy Questionnaire for Children

Therapy Outcome Questionnaire for Children

To be completed by the SLT (accompanied by audio/video recordings)
Questions 7-12 for children over 12 years of age

Name of Child: **Date:** **Treatment block:**

1. Which sounds did you work on in therapy?

   p b m f v t d n r s z sh l y k g ng h ch j w vowels

2. What did you think of using the iPad/looking at your tongue with the ultrasound (delete as appropriate)?

____
____
____
____
____
____
____
____
____

3. Did you enjoy the iPad/ultrasound sessions (delete as appropriate)? What was the best bit?

____
____
____
____
____
____
____
____
____

4. What was the worst bit or the hardest bit?

____
____
____
____
____
____
____
____
____

5. Were the sessions (circle):

   Too Short   Just Right   Too long

6. Do you think using the iPad/ultrasound (delete as appropriate) has helped your speaking? How has it helped?

____
____
____
____
____
____
____
____
____

7. Has anybody at home or at school noticed anything different about your speaking?

   Yes / No (delete as appropriate)

   If yes, describe what they have said:

____
____
____
____
____
____
____
____
____

8. How often do you think your parents/carer understands you when you speak?

Always Almost always Sometimes Rarely Never

9. How often do you think your brothers or sisters understands you when you speak? (if applicable)

Always Almost always Sometimes Rarely Never

10. How often do you think your teacher at school understands you when you speak?

Always Almost always Sometimes Rarely Never

11. How often do you think your friends understand you when you speak?

Always Almost always Sometimes Rarely Never

12. When you talk to new people, how often to they understand you when you speak?

Always Almost always Sometimes Rarely Never

13. Is there anything else you would like to say about taking part in the project?

____
____
____
____
____
____
____
____
____

7.4 Appendix 4: Parent Questionnaire

Therapy Outcome Questionnaire for Parents

Name of Child: **Date:** **Treatment block:**

1 Please complete the Intelligibility in Context Scale (provided separately)

2 Circle the speech sounds your child worked on:

p b m f v t d n r s z sh l y k g ng h ch j w vowels

3 Did your child work on (circle):

single words syllables whole words sentences key vocabulary

4 Did your child use (circle):

Ultrasound Speech Trainer 3D

5 Please rate your child's progress with their speech since enrolling in the project:

greatly improved moderately improved not improved slightly deteriorated greatly deteriorated

6 After treatment, my child's awareness of speech sounds has:

greatly improved moderately improved not improved slightly deteriorated greatly deteriorated

7 Please specify which speech sounds your child has learnt to say during therapy:

p b m f v t d n r s z sh l y k g ng h ch j w vowels

8 My child's ability to articulate speech sounds that s/he was not able to achieve before therapy has:

greatly improved moderately improved not improved slightly deteriorated greatly deteriorated

9 My child is able to use the new speech sounds in conversation:

always most of the time some of the time rarely never

10 Please rate your child's speech compared to siblings/peers without cleft palate or speech disorders:

much better slightly better the same slightly worse much worse

11 Please comment on whether or not you think using Speech Trainer 3D/Ultrasound (delete as appropriate) has made it easier for your child to achieve his speech therapy goals:

12 Any other comments?

7.5 Appendix 5: Three Month Post-Therapy Questionnaire for Children

3 Month Post-Therapy Outcome for Children

To be completed by the SLT

1. **Which tool did you like better? (circle picture)**

Ultrasound

iPad

2. **Why did you like it better?**

7.6 Appendix 6: Information Sheet for Perceptual Evaluation

Queen Margaret University

EDINBURGH

A Perceptual Experiment of Two Cleft Palate Speakers
Information Sheet for Potential Participants

You are being invited to take part in a research study. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve.

Please take time to read the following information carefully. Talk to others about the study if you wish. Contact us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of the study?
The aim of the overall research is to investigate the processes, designs and outcomes of visual tools used in assessment and speech therapy for children with a Cleft Palate, with a particular focus on Ultrasound Tongue Imaging (UTI) and an iPad app called Speech Trainer 3D. This is only one part of a bigger study being carried out at Queen Margaret University.
The main purpose of this part of the study is to investigate listener judgements on two children with repaired cleft palate who have received therapy using UTI and Speech Trainer 3D. We want to find out if each child's speech has improved after two blocks of therapy using an iPad app or ultrasound. The secondary aim is to investigate inter-rater reliability of listener judgements in assessment of speech in individuals with Cleft Palate. This study will require you listen to two different speakers on two different days. You will be given two different versions of words and will be asked to make a judgement on which version you think is closer to the target adult production.
Why have I been asked to take part?
You have been asked to take part because you are a speech and language therapy student at Queen Margaret University, in either level 3 or 4 of the BSc course or level 2 of the PgDip course. You have been asked to take part because you are phonetically trained listeners who should be able to make a clinical judgement on whether each child's speech has improved.

Do I have to take part?
No, it is up to you to decide whether or not you take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason.

What will happen if I take part?

• You will be asked to attend our Speech Science Laboratory at Queen Margaret University for two sessions.
• You will be asked to provide information on your level of study and whether or not you have previous experience with speech associated with Cleft Palate.
• During each session you will be asked to sit in front of a computer for approximately one hour where you will be asked to listen to some words through headphones.
• You will be asked to listen to data of two children with repaired cleft palate.
• Each question will provide two versions of the same word. You will be given a target word and will be asked to make a judgement on whether version A or B is closer to the target production of the word provided.
• This experiment will require you to attend twice, with approximately two months between each session. Each session should take no longer than 1 hour.
• All data will be anonymised and we will not use your name in any report or presentation.

What are the possible benefits of taking part?
If you take part, you will have the opportunity to listen to real-life clinical data of a client group that you may or may not have worked with on placement. At the end of the experiment you will be provided with feedback regarding your responses, if you request this.

What are the possible disadvantages and risks of taking part?
Due to the number of comparisons you will be asked to make, there is only a slight risk of tiredness. To overcome this, you will be provided with 6 rest breaks that you may or may not choose to take.

What happens when the study is finished?
If you would like us to, we will respond to you within one month of you completing the study with feedback regarding your responses and how you did, compared to other participants as a whole. No other participants will be given your personal scores.

Will taking part in the study be kept confidential?
All personal information (name etc) that we collect during the course of the research will be kept confidential and there are strict laws which safeguard your privacy at every stage. Data will be kept at Queen Margaret University, and anonymised.

What will happen to the results of the study?
The results of the study will be shared with the public, Speech and Language Therapists and academics via our website, conference presentations and publication in academic journals.

Who is organising the research and why?
This study has been organised by Queen Margaret University as part of a PhD studentship and is funded by Queen Margaret University. It will run from September 2011 to September 2014.

Who has reviewed the study?
The study proposal has been reviewed by independent reviewers for the funders. A favourable ethical opinion has been obtained from South East Scotland REC 01 and Queen Margaret University ethics committee. NHS management approval has also been obtained.

It is entirely up to you to decide whether or not to take part in the project. If you do decide to participate, you will be given this information sheet to keep and be asked to sign a consent form. You are free to withdraw from the study at any stage without giving a reason.

If you would like to consult an independent person, who knows about this project but is not involved in it, you are welcome to contact Dr Janet Beck, 0131 474 0000. You are also welcome to contact my PhD Director of Studies, Professor James Scobbie, on 0131 474 0000 (Say "Jim Scobbie" when prompted by the automated voice system) or email JScobbie@qmu.ac.uk.

If you have read and understood this information sheet, and you think you might be interested in participating in the study, please now email ZRoxburgh@qmu.ac.uk. There will be an opportunity to ask questions and sign the consent form when you come to Queen Margaret University to meet with the research SLT/ PhD Student, Zoe Roxburgh.

Thank you for taking the time to read this information.

Contact details of the researcher:
Name of PhD Student/ Speech and Language Therapist: Zoe Roxburgh

Address:
Speech and Hearing Sciences,
Queen Margaret University, Edinburgh
Queen Margaret University Drive
Musselburgh
East Lothian EH21 6UU
Email / Telephone: ZRoxburgh@qmu.ac.uk / 0131 474 0000

If you wish to make a complaint about the study, please contact NHS Lothian:

NHS Lothian Complaints Team
2nd Floor
Waverley Gate
2-4 Waterloo Place
Edinburgh
EH1 3EG
Tel: 0131 465 5708

7.7 Appendix 7: Consent Form for Perceptual Evaluation

Queen Margaret University

EDINBURGH

Participant Consent Form

Project Title: Visualising speech: the use of ultrasound in assessing speech sound disorders associated with cleft palate

Name of researcher: Zoe Roxburgh

Address: PhD Student, Speech and Hearing Sciences,
School of Health Sciences
Queen Margaret University, Edinburgh
Queen Margaret University Drive
Musselburgh
East Lothian EH21 6UU

Email / Telephone: ZRoxburgh@qmu.ac.uk / 0131 474 0000 (Say "Zoe Roxburgh" when prompted by the automated voice system)

Thank you for reading the information about our research project. If you would like to take part, please read and sign this form.

Participant's name ______________________________ Level of Study_________

Do you have any previous experience with cleft palate speech? Yes/No (Please Circle)

If yes, please report on what experience you have with this client group:

____
____
____
____

Please Initial Box

1. I have read and understand the information sheet dated 25/09/2013, (Version 1.0) and have had the opportunity to ask questions. ☐

2. I understand that participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my legal rights being affected.

3. I agree to participating in this study and that information collected via PRAAT can be stored and used indefinitely but anonymously for analysis, research, academic conference presentations, and future applications for research funding, and that the anonymous results of such analyses can be disseminated freely to audiences and research users of all types.

4. I agree to attend two sessions for 1 hour per session to complete two perception experiments.

5. I would like to be given feedback at the end of the experiment

Name of Participant	Signature	Date
Name of Person taking consent	Signature	Date
Researcher	Signature	Date

Thank you for agreeing to participate in this research.

7.8 Appendix 8: Slides Shown to Participants of Perceptual Evaluation

A Perceptual Experiment of Two Cleft Palate Speakers

Zoe Roxburgh
PhD Student

Aim of the study

- Investigate listener judgements on two children with repaired cleft palate who have received therapy using UTI and Speech Trainer 3D (iPad app)
- Find out if each child's speech has improved after two blocks of therapy using an iPad app or ultrasound
- Investigate inter-rater reliability of listener judgements in assessment of cleft palate speech

Benefits to you

- Opportunity to listen to real-life clinical data of a client group that you may or may not have worked with on placement
- Feedback regarding your responses
  - If requested
- Access to data that may be used for Masters projects

Disadvantages

- Slight risk of tiredness
  - provided with rest breaks

What we need you to do…

- Attend 2 sessions
  - Each lasting approx. 1 hour
- Listen to data of 2 speakers with repaired CP
- Each question will provide two versions of the same word. You will be given a target word and will be asked to make a judgement on whether version A or B is closer to the target production of the word provided

Speaker 1

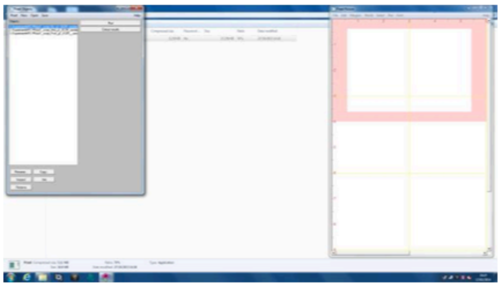
- Male
- Aged 9;2
- Repaired Submucuous Cleft Palate
- Treated /n/

Speaker 2

- Male
- Aged 6;3
- Repaired Submucuous Cleft Palate
- Treated velars /k g ŋ/

Instructions

- You will hear 108 word-pair trials in total, one pair at a time
  - Provided with rest breaks
- You will see one target word written on the screen
- You will be given two versions of the word to listen to
- You can listen to each pair up to 3 times



Which version spoken by the child is closer to the English target word written on the screen?

Please select either version A or version B.

7.9 Appendix 9: Example of MFC File used in PRAAT for the sub-study 2b (Craig)

```
"ooTextFile"
"ExperimentMFC 4"
stimuliAreSounds? <yes>
stimulusFileNameHead = "02ZR phase 2/"
stimulusFileNameTail = ".wav"
stimulusCarrierBefore = ""
stimulusCarrierAfter = ""
stimulusInitialSilenceDuration = 0.05 seconds
stimulusMedialSilenceDuration = 0 seconds
numberOfDifferentStimuli = 108

"angry_Mid1,silence0.5,angry_B2" "angry"
"banging_B2,silence0.5,banging_Mid1" "banging"
"bucket_Mid1,silence0.5,bucket_B2" "bucket"
"cage_Mid1,silence0.5,cage_B2" "cage"
"car_Mid1,silence0.5,car_B2" "car"
"carrots_B2,silence0.5,carrots_Mid1" "carrots"
"comb_Mid1,silence0.5,comb_B2" "comb"
"computer_B2,silence0.5,computer_Mid1" "computer"
"cookie_B2,silence0.5,cookie_Mid1" "cookie"
"cup_B2,silence0.5,cup_Mid1" "cup"
"flag_Mid1,silence0.5,flag_B2" "flag"
"gas_Mid1,silence0.5,gas_B2" "gas"
"gate_Mid1,silence0.5,gate_B2" "gate"
"goat_B2,silence0.5,goat_Mid1" "goat"
"gorilla_Mid1,silence0.5,gorilla_B2" "gorilla"
"guitar_Mid1,silence0.5,guitar_B2" "guitar"
"gum_B2,silence0.5,gum_Mid1" "gum"
"handbag_Mid1,silence0.5,handbag_B2" "handbag"
"jacket_B2,silence0.5,jacket_Mid1" "jacket"
"jog_Mid1,silence0.5,jog_B2" "jog"
"jumping_Mid1,silence0.5,jumping_B2" "jumping"
"kangaroo_B2,silence0.5,kangaroo_Mid1" "kangaroo"
"lego_Mid1,silence0.5,lego_B2" "lego"
"magic_B2,silence0.5,magic_Mid1" "magic"
"magnet_B2,silence0.5,magnet_Mid1" "magnet"
"necklace_B2,silence0.5,necklace_Mid1" "necklace"
"nuggets_Mid1,silence0.5,nuggets_B2" "nuggets"
"ring_B2,silence0.5,ring_Mid1" "ring"
"singer_Mid1,silence0.5,singer_B2" "singer"
"skiing_B2,silence0.5,skiing_Mid1" "skiing"
"smoke_B2,silence0.5,smoke_Mid1" "smoke"
"snack_B2,silence0.5,snack_Mid1" "snack"
"snowflake_Mid1,silence0.5,snowflake_B2" "snowflake"
"strong_Mid1,silence0.5,strong_B2" "strong"
"sugar_B2,silence0.5,sugar_Mid1" "sugar"
"warthog_B2,silence0.5,warthog_Mid1" "warthog"
```

"jog_mid2,silence0.5,jog_Post1" "jog"
"warthog_Post1,silence0.5,warthog_mid2" "warthog"
"ring_mid2,silence0.5,ring_Post1" "ring"
"handbag_Post1,silence0.5,handbag_mid2" "handbag"
"car_Post1,silence0.5,car_mid2" "car"
"cup_mid2,silence0.5,cup_Post1" "cup"
"bucket_Post1,silence0.5,bucket_mid2" "bucket"
"nuggets_mid2,silence0.5,nuggets_Post1" "nuggets"
"cage_mid2,silence0.5,cage_Post1" "cage"
"strong_mid2,silence0.5,strong_Post1" "strong"
"jumping_Post1,silence0.5,jumping_mid2" "jumping"
"kangaroo_Post1,silence0.5,kangaroo_mid2" "kangaroo"
"carrots_Post1,silence0.5,carrots_mid2" "carrots"
"smoke_mid2,silence0.5,smoke_Post1" "smoke"
"comb_Post1,silence0.5,comb_mid2" "comb"
"gum_mid2,silence0.5,gum_Post1" "gum"
"sugar_mid2,silence0.5,sugar_Post1" "sugar"
"magnet_mid2,silence0.5,magnet_Post1" "magnet"
"gate_mid2,silence0.5,gate_Post1" "gate"
"jacket_mid2,silence0.5,jacket_Post1" "jacket"
"guitar_mid2,silence0.5,guitar_Post1" "guitar"
"magic_Post1,silence0.5,magic_mid2" "magic"
"banging_Post1,silence0.5,banging_mid2" "banging"
"lego_Post1,silence0.5,lego_mid2" "lego"
"gorilla_Post1,silence0.5,gorilla_mid2" "gorilla"
"skiing_mid2,silence0.5,skiing_Post1" "skiing"
"gas_Post1,silence0.5,gas_mid2" "gas"
"snowflake_Post1,silence0.5,snowflake_mid2" "snowflake"
"angry_mid2,silence0.5,angry_Post1" "angry"
"snack_Post1,silence0.5,snack_mid2" "snack"
"flag_mid2,silence0.5,flag_Post1" "flag"
"singer_mid2,silence0.5,singer_Post1" "singer"
"computer_Post1,silence0.5,computer_mid2" "computer"
"necklace_mid2,silence0.5,necklace_Post1" "necklace"
"goat_Post1,silence0.5,goat_mid2" "goat"
"cookie_Post1,silence0.5,cookie_mid2" "cookie"
"cage_post2,silence0.5,cage_B1" "cage"
"singer_post2,silence0.5,singer_B1" "singer"
"jog_post2,silence0.5,jog_B1" "jog"
"strong_post2,silence0.5,strong_B1" "strong"
"magnet_post2,silence0.5,magnet_B1" "magnet"
"sugar_B1,silence0.5,sugar_post2" "sugar"
"car_post2,silence0.5,car_B1" "car"
"snowflake_post2,silence0.5,snowflake_B1" "snowflake"
"skiing_post2,silence0.5,skiing_B1" "skiing"
"gum_post2,silence0.5,gum_B1" "gum"
"smoke_post2,silence0.5,smoke_B1" "smoke"
"guitar_B1,silence0.5,guitar_post2" "guitar"
"jacket_post2,silence0.5,jacket_B1" "jacket"
"snack_B1,silence0.5,snack_post2" "snack"
"flag_post2,silence0.5,flag_B1" "flag"
"jumping_B1,silence0.5,jumping_post2" "jumping"
"gate_B1,silence0.5,gate_post2" "gate"

```
"bucket_post2,silence0.5,bucket_B1" "bucket"
"gorilla_B1,silence0.5,gorilla_post2" "gorilla"
"ring_B1,silence0.5,ring_post2" "ring"
"kangaroo_post2,silence0.5,kangaroo_B1" "kangaroo"
"cup_post2,silence0.5,cup_B1" "cup"
"nuggets_post2,silence0.5,nuggets_B1" "nuggets"
"handbag_B1,silence0.5,handbag_post2" "handbag"
"banging_post2,silence0.5,banging_B1" "banging"
"comb_post2,silence0.5,comb_B1" "comb"
"lego_B1,silence0.5,lego_post2" "lego"
"angry_B1,silence0.5,angry_post2" "angry"
"warthog_B1,silence0.5,warthog_post2" "warthog"
"cookie_B1,silence0.5,cookie_post2" "cookie"
"computer_B1,silence0.5,computer_post2" "computer"
"necklace_B1,silence0.5,necklace_post2" "necklace"
"magic_B1,silence0.5,magic_post2" "magic"
"goat_B1,silence0.5,goat_post2" "goat"
"gas_B1,silence0.5,gas_post2" "gas"
"carrots_B1,silence0.5,carrots_post2" "carrots"

numberOfReplicationsPerStimulus 1
breakAfterEvery = 18
randomize = <PermuteBalancedNoDoublets>
startText = "Click to start the experiment."
"Which utterance is a better version of the word at the top - first or second?"
pauseText = "Break?"
endText = "The experiment has finished."
maximumNumberOfReplays = 2
replayButton = 0.3 0.7 0.1 0.2 "Play again?" ""
okButton = 0.7 0.95 0.05 0.15 "OK, go on to next pair" ""
oopsButton = 0 0 0 0 "" ""
responsesAreSounds? <no> "" "" "" "" 0 0
numberOfDifferentResponses = 2
  0.3 0.5 0.7 0.85 "first" "150" "first"
  0.5 0.7 0.7 0.85 "second" "150" "second"

numberOfGoodnessCategories = 5
  0.05 0.35 0.40 0.50 "1 (just guessing)"
  0.35 0.45 0.40 0.50 "2"
  0.45 0.55 0.40 0.50 "3"
  0.55 0.65 0.40 0.50 "4"
  0.65 0.95 0.40 0.50 "5 (very sure indeed)"
```

7.10 Appendix 10: Published Journal Article

ROXBURGH, Z., CLELAND, J. and SCOBBIE, J.M., 2016. Multiple phonetically trained-listener comparisons of speech before and after articulatory intervention in two children with repaired submucous cleft palate. *Clinical Linguistics & Phonetics*. vol. 30, no. 3-5, pp. 398-415. Available from: http://dx.doi.org/10.3109/02699206.2015.1135477

PDF attached separately

7.11 Appendix 11: List of Abbreviations

A
AAA – Articulate Assistant Advance
AB – Earlier/Later
AoS – Apraxia of Speech
App – Application
Ax – Assessment

B
BA – Later/Earlier
BCLP – Bilateral Cleft Lip and Palate
BL - Baseline
BPVSIII – British Picture Vocabulary Scale Third Edition

C
CAS – Childhood Apraxia of Speech
CASL – Clinical Audiology, Speech and Language
CCH – Community Child Health
CELF-4 – Clinical Evaluation of Language Fundamentals Fourth Edition
CL – Cleft Lip
CI – Confidence Interval
CLP – Cleft Lip and Palate
CL +/- P – Cleft Lip with or without Palate
Conf - Confidence
CP – Cleft Palate
CPO – Cleft Palate Only
CTR – Cleft Type Realisation
CV – consonant-vowel
CVC – consonant-vowel-consonant

D
DEAP – Diagnostic Evaluation of Articulation and Phonology
DEI – Dorsum Excursion Index
DSSD – Developmental Speech Sound Disorder

E
EDS – Eating, Drinking and Swallowing
EMA – Eletcromagnetic Articulography
ENT – Ear, Nose and Throat
EPG – Electropalatography
ExtIPA – Extended International Phonetic Alphabet

G
GOS.SP.ASS – Great Ormond Street Speech Assessment

H
HI – Hearing Impairment

I
ICS – Intelligibility in Context Scale
IPA – International Phonetic Alphabet / Association
IRAS – Integrated Research Application System

K
KP – Knowledge of Performance
KR – Knowledge of Results

M
M/Maint – Maintenance
MDT – Multi-Disciplinary Team
MFC – Multiple Forced Choice
MRI – Magnetic Resonance Imaging

N
NHS – National Health Service

P
PCC – Percent Consonant Correct
PPSA - Phonetic and Phonological Systems Analysis
PTCC – Percent Target Consonant Correct

Q
QMU – Queen Margaret University

R
R&D – Research and Development
RCSLT – Royal College of Speech and Language Therapists
RT – Reaction Time

S
SFWF – Syllable Final, Word Final
SIWI – Syllable Initial, Word Initial
SMCP – Submucous Cleft Palate
SLT – Speech and Language Therapy / Therapists
SS-ANOVA – Smoothing Spline ANOVA
SSD – Speech Sound Disorder

T
TCPI – Tongue Constraint Position Index TOM – Therapy Outcome Measures
tSLT – treating Speech and Language Therapist
Tx - Treatment

U
UCLP – Unilateral Cleft Lip and Palate
UK – United Kingdom

UTI – Ultrasound Tongue Imaging
UVBF – Ultrasound Visual Biofeedback

V
V1 – Version One
V2 – Version Two
VAM – Visual Articulatory Model
VAS – Visual Analogue Scale
VBF – Visual Biofeedback
VC – vowel-consonant
VCV – vowel-consonant-vowel
VF - Videofluoroscopy
VP – Velopharyngeal
VPD – Velophayngreal Dysfunction

W
WF – Word Final
WI – Word Initial
WM – Word Medial

#
2D – Two Dimensional
3D – Three Dimensional